Aberystwyth University

Doctoral Thesis

Computer-Aided Detection of Prostate Cancer within the Peripheral Zone in T2-Weighted Magnetic Resonance Imaging

Yambu Andrik Rampun

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

in the

Department of Computer Science

March 2016

Declaration of Authorship

DECLARATION

This work has not previously been accepted in substance for any degree and is not being currently submitted in candidature for any degree

Signed: _____

Date: _____

STATEMENT 1

This thesis is the result of my own investigations, except where otherwise stated. Other sources are acknowledged by footnotes giving explicit references, A bibliography is appended.

Signed: _____

Date: _____

STATEMENT 2

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and the title and summary to be made available to the outside organisations.

Signed: _____

Date: _____

"The prime quality of a successful PhD student is DILIGENCE, not INTELLIGENCE!....
....The rest is hard work and inspiration"

Unknown

Abstract

The aim of this research is to develop a computer-aided detection system for prostate cancer within the peripheral zone, based on single modality T2-W Magnetic Resonance Imaging. As the most diagnosed and second leading cause of death from cancer in men, prostate cancer is a significant health problem globally. In fact, considering the deficiencies in current clinical screening methods, there is a need for a rapid development of MRI technologies and computer algorithms to better detect prostate cancer.

In this thesis, we developed one unsupervised, and two supervised computer algorithms, using different texture descriptors involving different resolutions, filters, techniques and orientations. For classification purposes, we investigated 11 machine learning algorithms to build predictive models. This thesis also investigated the effects of window sizes for all performance metrics. In the proposed unsupervised method a small number of texture descriptors were used to differentiate benign and malignant regions. Subsequently, the fuzzy c-means clustering algorithm was employed to segment malignant regions. The resulting binary segmentations were then combined to find overlapping regions with the highest probability of being malignant. In contrast, the two supervised methods were based on 215 texture descriptors and textons. Both methods used the same machine learning algorithms in the training and classification phases. To evaluate the performance of the proposed methods, the unsupervised and supervised methods were tested based on 37 (275 MRI images) and 45 (418 MRI images) patients, respectively. The performance evaluations showed that the results of all methods were comparable with the state-of-the-art in the literature in terms of area under the curve, classification accuracy, sensitivity and specificity.

The contributions of this thesis are three-fold. First, we developed novel supervised and unsupervised computer-aided detection systems for prostate cancer, which are different from those produced hitherto, in the sense of our methods exploit different combinations of texture descriptors and machine learning algorithms. In fact, this thesis is the first study which has thoroughly investigated textons in prostate cancer detection. Secondly, this thesis made an extensive investigation into the effects of window size, both on the methods and feature's performance itself. Thirdly, we provide an extensive quantitative comparison of the performances of 11 machine learning algorithms and a thorough qualitative comparison between our results and the state-of-the-art in the literature.

Acknowledgements

Thank you God for your abundant blessings which have given me resilience in the face of life's challenges. First and foremost, I would like to thank my supervisor Prof. Reyer Zwiggelaar for his continuous guidance and encouragement throughout the course of my PhD. He has been a tremendous teacher through the various stages of my research. Without his insight, direction and suggestions I would have not been able to produce this thesis. His scientific knowledge has been a light along the journey of my PhD. My special thanks also go to Dr. Bernie Tiddeman for his helpful suggestions and comments, which helped to shape the final version of my journal papers and thesis.

I would also like to mention the following people for their assistance, which significantly helped my research; Dr. Zhili Chen, Dr. Harry Strange, Raquel Gil and Ling Zheng. To all members of the Vision, Graphics and Visualisation Group, thanks for their feedback, questions, discussions, and suggestions for my research. My acknowledgement must also go to the person who first introduced me to research, Dr. Trevor Barker; without his encouragement and suggestions, I would have never started a PhD. Similarly I must acknowledge those who selected me for a Doctoral Career Development Scholarship (DCDS) which has given me invaluable financial support for my research.

My acknowledgement will never be complete without expressing my deepest gratitude to my family, especially my parents who have been giving me lots of love, support and blessings. I also wish to thank my friends Dr. Robert Farmer and Barry Draper for their constant motivation in facing difficult times during my PhD. I am so grateful to have such family and friends who are always there in times of my need.

Contents

List of Figures

List of Tables

Abbreviations

ADC	Apparent Diffusion Coefficient
ADTree	Alternating Decision Tree
ANN	Artificial Neural Network
$\boldsymbol{A_z}$	Area Under the Curve
BNet	Bayesian Network
C4.5	J48
$\boldsymbol{CA}$	Classification Accuracy
CAD	Computer Aided Detection
CAD-PC	Computer Aided Detection system for Prostate Cancer
CSLBP	Center-Symmetric Local Binary Pattern
CT	Computed Tomography
CZ	Central Zone
DCE	Dynamic Contrast Enhanced
DRE	Digital Rectal Examination
DW	Diffusion Weighted
DWT	Discrete Wavelet Transform
ELBP	Elongated Local Binary Pattern
FCM	Fuzzy C-Mean
GLCM	Grey Level Co-occurrence Matrix
GLDM	Grey Level Difference Matrix
GLRLM	Gray Level Run Length Matrix
GLCM	Gray Level Co-occurrence Matrix
HOG	Histogram Orientation Gradient
$\boldsymbol{k}$**-NN**	k- Nearest Neighbour

LBP	Local Binary Pattern
LDA	Linear Discriminant Analysis
LoG	Laplacian of Gaussian
LOOCV	Leave-One Out Cross-Validation
LTP	Local Ternary Pattern
MBP	Median Binary Pattern
MDL	Minimum Description Length
MI	Mutual Information
MLP	MultiLayer Perceptron
MRI	Magnetic Resonance Imaging
MRS	Magnetic Resonance Spectroscopy
M-V (b2)	Ensambled best 2 classifiers
M-V (b3)	Ensambled best 3 classifiers
NB	Naïve Bayes
***n*-FV**	n-fold Cross-Validation
PCA	Principal Component Analysis
PDF	Probability Density Functions
PET	Positron Emission Tomography
PSA	Prostate-Specific Antigen
PV	Prostate Volume
PZ	Peripheral Zone
RBF	Radial Basis Function
ROC	Receiver Operating Characteristic
ROI	Region Of Interest
RF	Random Forests
Sen	Sensitivity
SGLDM	Spatial Gray-Level Dependence Matrix
Spe	Specificity
SL	Simple Logistic
SVM	Support Vector Machine
T	Tumour

T2-W	T2- Weighted
TRUS	Transrectal Ultrasound Guided Biopsy

Chapter 1

Introduction

The aim of this research is to develop Computer Aided Detection methods for prostate cancer detection within the peripheral zone in T2-Weighted Magnetic Resonance Imaging (MRI). This is done by developing a variety of computer algorithms employing texture analysis to distinguish benign and malignant tissues. In this chapter prostate cancer facts and figures will be presented covering the world, Europe and the United Kingdom statistics which include incidence and mortality rates. The next section of this chapter then reviews the main clinical diagnostic methods for prostate cancer and their deficiencies. Subsequently, Magnetic Resonance Imaging technology will be briefly reviewed in prostate imaging and its challenges in the field of medicine. This includes brief explanations of how computers can be used in helping radiologists to speed up medical image interpretation, hence assisting radiologists in diagnostic decision making. Finally, the scope and objectives of this thesis will be presented and an outline of this thesis is provided in the final section of this chapter.

1.1 Prostate Cancer Facts and Figures

Prostate cancer is the most commonly diagnosed cancer amongst men, and remains the second leading cause of deaths from cancer in men. The causes of prostate cancer remain uncertain. However, studies [3, 4] suggested that prostate cancer is more common among men over 50 years old, those with a family history of the disease, those from particular ethnic/genetic groups, those with excess body weight, and those who have low levels of physical activity. According to the latest world cancer statistics published by the American Cancer Society [5] prostate cancer is the fourth most

common cancer globally and is surpassed only by lung, female breast, and bowel cancers. In 2013 the number of cases diagnosed in the UK and USA were around 40,000 and 237,000 respectively. The global figure is predicted to reach 1.7 million cases by 2030 in comparison to the 1.1 million in 2012 [6–8]. Statistically, nine out of ten men survive for at least five years if the cancer is diagnosed at its earliest stage [9]. However, early definitive diagnosis of prostate cancer remains problematic due to the lack of a simple unambiguous test. The next section will briefly present the statistics regarding the prostate cancer incidence and mortality per 100,000 men relative to the World, Europe and the UK, data compiled by the International Agency for Research on Cancer of the World Health Organisation (WHO) and Cancer Research UK in 2012 [10].

1.1.1 World Statistics

Prostate cancer is one of the most common cancers affecting men, and is a major killer, with an estimated 1.1 million diagnoses and 307,000 deaths in 2012, the last year for which comprehensive data are available from [10]. Figure 1.1 shows striking features in the worldwide data for incidence and mortality. In the 'more developed regions' the incidence per 100,000 men exceeds 60, peaking at more than 110 in Australia/New Zealand, but is relatively low at around 35 in Eastern and Central Europe. Northern America and western Europe came second and third with more than 95 incidences. Most Asian countries show low incidence rates; generally below

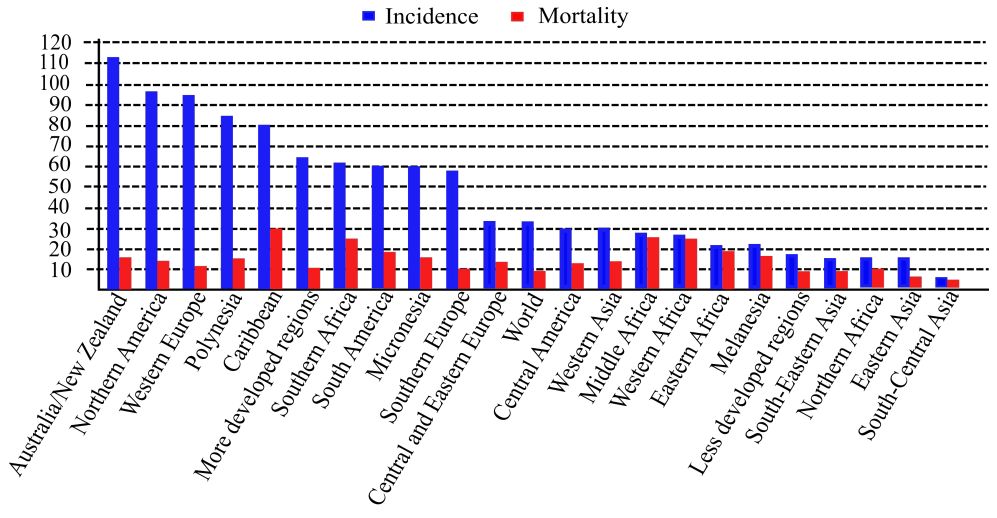

FIGURE 1.1: Prostate cancer incidence and mortality statistics per 100,000 men across the globe in 2012 [10].

30. Much of the variability in apparent incidence seems certain to be due to varying standards of diagnosis and detection. It also seems certain that the real incidence is rather less variable than shown in Figure 1.1. In contrast with incidence, mortality rates worldwide are much less variable. The data do suggest that Carribean and some African nations have the highest mortality rates with more than 20 men dying of prostate cancer per 100,000 men, whilst some Asian regions have the lowest mortality rates. Although the incidence rates from more developed countries are at least triple the incidence rates from less developed countries, both have very similar mortality rates of around 20. This maybe caused by better screening and treatment methods compared to those in less developed countries.

1.1.2 Europe Statistics

In Europe, there were around 400,000 cases reported in 2012 and the UK came 3rd highest in terms of the number of incidence with around 45,000 cases and came 16th based on the number of incidence rates per 100,000. Figure 1.2 shows the data with respect to Europe with a similar pattern of greater variability in incidence than mortality that was seen in the global data. Incidence is highest in Northern Europe in countries such as Norway (193.2), Sweden (175.2), Iceland (168.7), Switzerland (158.6), Estonia (145.4) and Finland (145.2), which also have the high mortality rates. The mortality rates are relatively low in Italy (14.1), Spain (15.2) and Malta (13.6) and some countries of south eastern Europe such as Albania (13.4), Moldova (15.6), Ukraine (15.7),

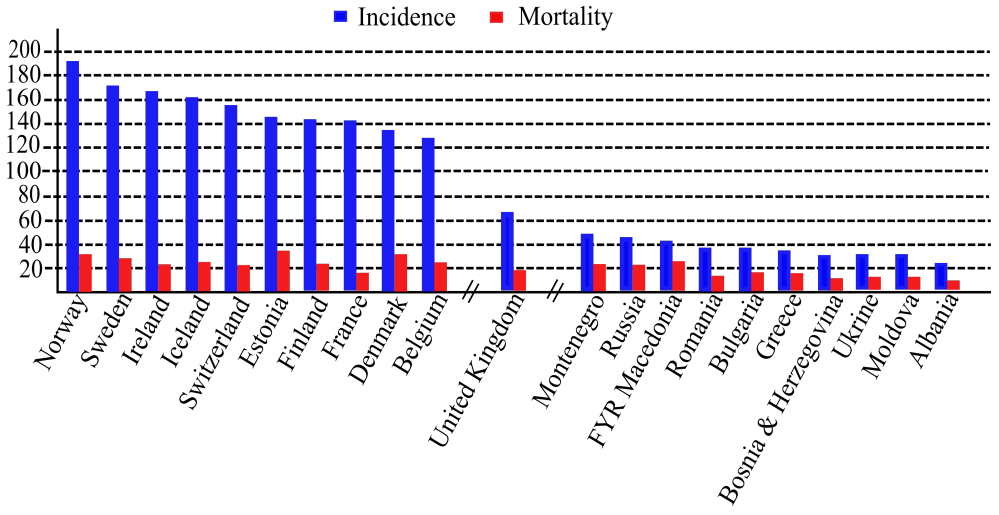

FIGURE 1.2: Ten countries in Europe with the highest and lowest number of prostate cancer incidence per 100,000 men [10].

Bosnia Herzegovina (15.1), Greece (17.7), Bulgaria (17.1) and Romania (16.3), which also have relatively low apparent incidences. The UK has slightly higher than average incidence and mortality rates, at around 70 and 20, respectively.

1.1.3 The United Kingdom Statistics

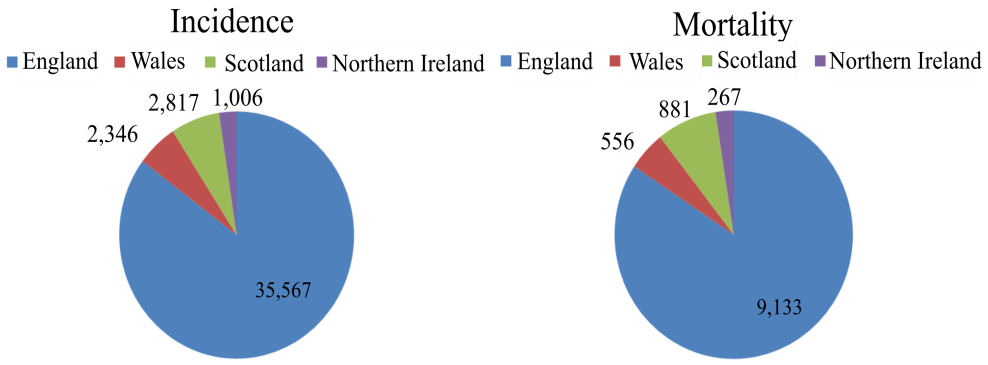

FIGURE 1.3: Prostate cancer incidence and mortality statistics in the UK in 2012 [11].

In 2012, there were 41,736 incidence reported in the UK with 10,837 deaths. Prostate cancer rates in the UK have at least tripled over the last 35 years causing prostate cancer to be the most common cancer among British men. Chart pies in Figure 1.3 show the incidence and mortality in the UK with England having the highest incidence with more than 35,000 cases. Scotland and Wales have similar numbers between 2,000 to 3,000 cases where Northern Ireland has the lowest incidence of around 1,000 cases. In terms of mortality rates, England remains the highest with approximately 9,100 cases ahead of Scotland and Wales with 881 and 556 deaths, respectively. Northern Ireland has the lowest mortality with 267 deaths. Taking account of population size the incidences and mortalities in Wales are both higher than in Scotland and England.

1.2 Prostate Anatomy

The prostate gland is part of a man's reproductive and urinary systems. The prostate is oval shaped with a rounded tip and approximately $4cm$ wide and $3cm$ thick [15]. Figure 1.4 shows the location of the prostate within man's reproductive and urinary

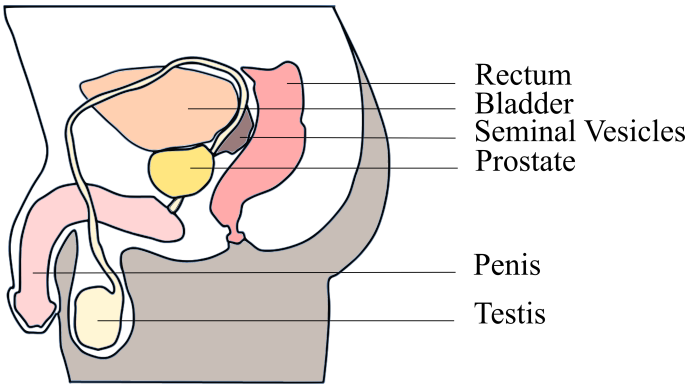

FIGURE 1.4: The location of prostate gland in human body anatomy adapted from [12–14].

systems which is just in front of the rectum, surrounding the base (or neck) of the bladder; it has 2 lobes that surround the urethra [12, 15, 16]. A normal human male prostate is usually slightly larger than a walnut and mostly contains glandular tissues (smooth tissues) [17]. Figure 1.5 shows 3D and 2D schematic models of the prostate adapted from [12–14]. The prostate gland is divided into three main zones namely peripheral zone (PZ), central zone (CZ) and transition zone (TZ). There are some other minor regions within the prostate such as fibromuscular zone, periurethral gland region and transition zone. However, these regions are less effected by tumours. Most prostate cancers start to develop within the PZ [18]. According to [12, 15, 16], approximately 75% to 80% of prostate cancers arise within the PZ and prostate cancer that arises within the this region is more aggressive than that which arises in the CZ.

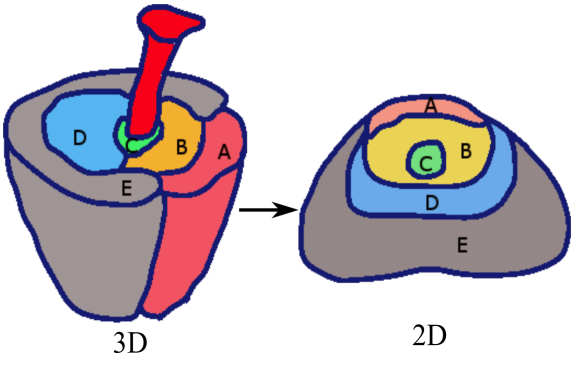

FIGURE 1.5: Prostate anatomy adapted from [12–14]: (a) fibromuscular zone, (b) transition zone, (c) periurethral gland region, (d) central zone and (e) peripheral zone.

1.3 Prostate Cancer Detection Methods

Over the last decade, several different methods have been introduced to detect prostate cancer. However, only a small number of methods is currently available in many hospitals. The two most common methods used for preliminary screening are the prostate-specific antigen (PSA) test and digital rectal examination (DRE). Transrectal ultrasound (TRUS) guided biopsy which is a more reliable method is often used for further screening.

1.3.1 Prostate-Specific Antigen

In many hospitals, PSA is the first screening procedure for men who are aged over 50. The PSA test is a blood test that measures the amount of antigen protein produced in blood. It is normal for all men to have a small amount of PSA in their blood, and this amount rises as men get older [19]. In general, if the patient's PSA level has risen to more than the normal level for his age, it could be a sign of prostate cancer. For example, according to the guide given by Prostate Cancer UK a normal PSA level for men aged 50-59 is $3ng/ml$. Normal PSA levels for men aged 60-69 and ≥ 70 are $4ng/ml$ and $5ng/ml$, respectively [19] (note that this guideline might be different in some hospitals). However, an elevated PSA level does not always indicate occurrence of prostate cancer because several factors can increase PSA level such as a urine infection, vigorous exercise and ejaculation in the 48 hours before a PSA test [19]. According to Schroder *et al.* [20], although the use of PSA reduces the rate of death by 20%, the benefit was associated with a high risk of overdiagnosis and overtreatment. In addition, the PSA test is not able to predict the aggressiveness of cancer. As a result, slow-growing and non-aggressive prostate cancer is frequently diagnosed in older patients [21].

In terms of accuracy, according to the study in [22] for every 100 men over 50 who have the PSA test, 10 of them have elevated PSA level and 3 of these are found to have prostate cancer. On the other hand, 88 men out of the other 90 men with a normal PSA level did not have prostate cancer. These figures show that PSA tests have low rates in specificity. The PSA test remains the most common preliminary diagnosis test as it has been demonstrated that an early diagnosis is best achieved using a combination of DRE and PSA [23].

1.3.2 Digital Rectal Examination

DRE is the cheapest and easiest screening method to detect prostate cancer because the doctor can examine the patient without the use of additional tools such as needles, ultrasound, microscope and patients do not have to make any preparations before the test. This test is usually performed after PSA test because DRE test itself can increase PSA level [19]. During the examination, the doctor gently inserts a lubricated, gloved finger into the patient's rectum to feel any sign of abnormalities such as lumps, soft or hard spots, etc. [24] (healthy prostate has smooth surface). If the size of patient's prostate is larger than normal, then a further investigation needs to be done. In addition, a prostate with a hard bumpy area could also be a sign of abnormality.

DRE test is not very accurate as a screening method partly due to experience variability among examiners. For example, a more experienced examiner can detect subtle abnormalities in comparison to less experienced clinicians. Moreover, small tumours are more difficult to detect than larger ones and if a tumour is located away from the rectal wall, clinicians will not able to palpate it during the examination [25]. A study conducted by Palmerola *et al.* [25] revealed that only 136 men were diagnosed with prostate cancer out of 290 patients identified as having an abnormal prostate gland. Another study of 6,630 men volunteering for DRE and PSA tests conducted by Catolana *et al.* [26] found that 45% of the cancers that were detected were having normal DRE. Palmerola *et al.* [25] concluded that an abnormal DRE had 0.44 sensitivity and 0.68 specificity. Nevertheless, this method remains a standard procedure for initial screening for prostate cancer as it is the easiest and quickest method.

1.3.3 Transrectal Ultrasound Guided Biopsy

The TRUS biopsy test is currently the most accurate method in comparison to PSA and DRE tests [23]. This test involves using thin needles (also known as a biopsy gun) to take around 10 to 12 sample tissues from the prostate gland [27]. The doctor uses a lubricated (to ease discomfort) ultrasound probe and put it into the patient's rectum (via the back passage). The ultrasound will scan the prostate's surface and the doctor uses the ultrasound image as a guide to capture sample tissues using the needles [28]. Subsequently, the sampled tissues will be analysed under the microscope by a pathologist. The accuracy of the results depend on the tissue samples from the patient's prostate gland. The main disadvantage of this method is that the needle

used to capture samples can miss cancerous tissues during the procedure, meaning the TRUS biopsy test result will show a negative result.

A recent study by Lacetera *et al.* [26] found that only 32 patients had prostate cancer out of 66 suspect undergoing TRUS examination. Moreover, studies in [21, 27] showed that systematic biopsies do not detect all clinically significant cancers and nearly a quarter (23%) of detectable cancers were missed. Daneshgari *et al.* [28] who used computerised biopsy simulations showed that the chance of missing a cancer by sextant biopsy is estimated at approximately 25%. Moreover, Svetec *et al.* [29] performed an ex vivo sextant biopsy on 90 prostates removed for biopsy-proven cancer, found that 41 prostates (46%) were negatives using TRUS biopsy whereas Norberg *et al.* [30] reported that the TRUS biopsy test could produce up to 35% false negatives. Nevertheless, this method is the most reliable method at present despite its shortcomings [23].

1.4 Magnetic Resonance Imaging (MRI)

In general, MRI is a non-invasive imaging technique that uses a strong magnetic field and pulses of radio wave energy to create the structural images of the organ [31, 32]. MRI technology was first used in 1977 and has become successful and widely used in many different applications. In radiology, MRI plays an important role for radiologists to non-invasively investigate internal parts of human body.

1.4.1 MRI Modalities

There are many different types of MRIs [31], however the most popular MRI modalities are: T1-Weighted (T1-W), T2-Weighted (T2-W), dynamic contrast-enhanced (DCE), magnetic resonance spectroscopic (MRS) and diffusion MRI (DWI). T1-W and T2-W MRIs are the most common modalities in radiology, DWI is one of the latest modalities and DCE and MRS are popular due to their ability to provide additional information not available in conventional MRIs (e.g. T1-W and T2-W MRI). T1-W MR images are produced when the machine is programmed to capture the longitudinal movement of protons and is usually used to look at normal anatomical details [33, 34]. On the other hand, T2-W images map the transverse movement of protons and are usually used to find diseases because unhealthy tissues tend to have a

higher water content than normal. T1-W and T2-W MRI are usually done in a 1.5T (Tesla) and in some cases 3.0T magnetic field strength.

In comparison to the conventional MRIs (T1-W and T2-W), DCE can give information about physiological tissue characteristics. In general, DCE uses T1-W scan for a fundamental imaging and uses MRI contrast agent called Gadolinium (which is injected into the patient's body) to produce MR images with higher contrast [35, 36]. On the other hand, MRS differentiate the metabolites composition (e.g. Choline and Creatine) of normal and abnormal tissues by analysing molecules such as hydrogen ions or protons [37]. The frequency of these metabolites are plotted on a graph as peaks of varying height and abnormality can be determined by analysing the graph. DWI is a new imaging technology which maps the motion of the water molecules within the organ tissues [38]. Note that there are other non-MRI scanning procedures such as ultrasound, positron emission tomography (PET) and X-ray computed tomography (X-ray CT).

Recently, the fusions of different modalities to analyse tissues abnormalities are becoming popular due to the ability to provide additional information (e.g. metabolic changes and pharmacokinetics parameters such as (wash-in rate K^{trans} and wash-out rate K_{ep}). In comparison, single modality is only able to provide limited information about the anatomical structures of the organ. For example, by comparing corresponding pixel values of T1-W and DCE images it is possible to identify permeable blood vessels and tumor tissue [36] which is not possible in T1-W or T2-W alone. In brain imaging, multiparametric MRI has been shown to improve the accuracy over single modality [39]. For example in acute stroke patients, the map risk can be generated by combining DWI and T2-W MRI on a voxel-wise basis to find the voxels with high probability of being malignant. In addition, multiparametric MRI also allows volumetric and signal analysis by performing statistical tests [39]. In liver imaging, T1-W MRI can provide a good intensity map to differentiate abnormal regions. However, DCE is needed to increase intensity appearance of the arterial and portal venous and T2-W is used to demonstrate the normal intrahepatic biliary structures [40].

On the other hand, in lung imaging [41] combining PET and conventional MRI seems to be advantageous due to higher soft tissue contrast [42] in comparison to MRI alone due to limitations such as low proton density in the lung (hence poor MR images) and sensitivity of movement [43]. Although multiparametric MRI is developing and holding promise in basic medical research, its development further remains very challenging both from the engineering and clinical points of view. For

example the first commercial PET/MRI prototype for a human scale hybrid scanner was not unveiled until 2007 [44]. However the, lack of standardised diagnostic criteria make the interpretation of multiparametric data heavily dependent on radiologist's experience. Moreover, the huge amount of image data that must be analysed in each case can be computationally expensive and time consuming [45].

1.4.2 Prostate Imaging in MRI

Since the reliability of clinical methods in screening and detecting prostate cancer is still questionable, integrating MRI into clinical practices (e.g. MRI/Ultrasound guided biopsy and multimodality image fusion) is becoming popular as it has shown a significant improvement over PSA and TRUS alone [20, 46]. Unfortunately such methods require substantial expertise from the radiologist, a significant amount of human interaction which increases the potential of human errors, time consuming and often suffer from observer variability [20, 46]. Therefore, in most hospitals, MRI is used to identify the staging of prostate cancer and treatment planning rather than to detect prostate cancer. In prostate MRI screening, a patient will lie on a table and move slowly into the scanner. The scanning process takes between 30 to 45 minutes.

MRI has been successful in brain, breast, lung and bone imaging [31]. Nowadays, MRI plays an important role in investigation of the anatomical human body as it is non-invasive, faster and has fewer complications. In prostate imaging, CAD system can help radiologists to identify abnormality by delineating regions with the highest probability of being malignant. The outlined regions (segmentation results) are useful as they provide a 'second opinion' for the radiologist in making decision whether a cancer is truly present or not.

Figure 1.6 shows an example of prostate T2-W MRI and a zoomed prostate gland with its ground truth delineated by an expert radiologist. Visually, we can see that the appearance of the malignant region is darker (lower intensity) compared to the healthy region within the PZ. Similarly, studies [12, 16, 47, 48] suggested that most cancer regions in the PZ tend to have a dark appearance on a T2-W MRI image. In fact, radiologists tend to use darker regions as the basis of their *priori* knowledge to identify abnormality within the PZ [49]. Nevertheless, in some cases low intensity does not always represent malignancy (resulting in false positive detection) due to inflammation and post-biopsy scarring [50]. Certain conditions such as benign prostatitis and intra-prostatic hemorrhage can cause similar appearances, therefore studies have reported

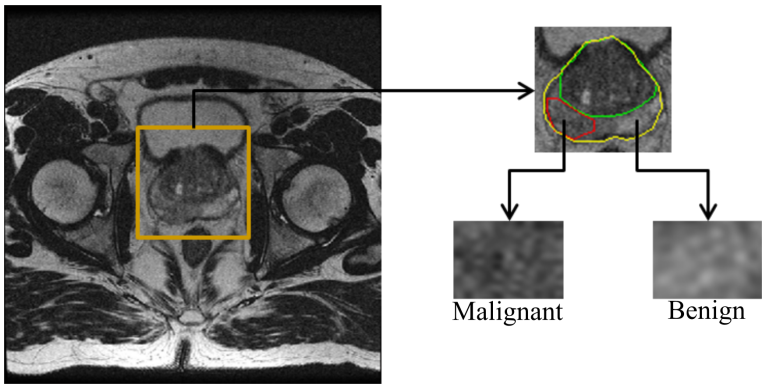

FIGURE 1.6: Left shows an example of Prostate T2-W MRI and upper right image is zoomed prostate gland with ground truth (yellow:prostate gland, green: CZ and red: tumour). The lower right shows samples of zoomed malignant and benign images.

that T2-W MRI has a low specificity and sensitivity ranging between 54%-82% and 46%-92%, respectively in detecting prostate cancer [51–53].

As with the other parts of human body, the use of multiparametric MRI for the prostate is becoming popular due to additional information which can be extracted from different modalities to improve the accuracy of diagnosis. Engelbrecht *et al.* [54] reviewed the use of multiparametric MRI and concluded that it can be used to assess the whole gland including the challenging anterior part which is the CZ. The main reason for using multiparametric MRI is due to insufficient information being provided by an individual modality and the amount of noise particularly in T1-W and T2-W MR images can significantly influence the image quality. For example although T2-W MR images can provide better structural details of the tissues in comparison to MRS and DCE they do not provide other information such as metabolic changes and pharmacokinetics parameters which can be very useful to support radiologist's decision whether a tumour is truly present or not.

In assessing radiologist's performance using mono and multiparametric MRI, Sciarra *et al.* [55] conducted a study using DCE and MRS in prostate cancer detection and found that combining these modalities resulted 87% sensitivity in comparison to 84% (DCE alone) and 71% (MRS alone). Apparent diffusion coefficient (ADC) which can be extracted from DWI was shown to improve the diagnostic performance by at least 15% sensitivity in comparison to T2-W MRI alone [56]. The ADC is a numerical map representation of magnitude of diffusion (of water molecules) within tissues. Malignant prostate tissues tend to have significantly lower values compared

to benign tissues [54]. The tissue metabolism in MRS shows that there is reduced level of citrate and elevated choline levels, compared with healthy prostate tissue [54]. Therefore, combining these data with other *priori* knowledge such as intensity and shape information from T1-W and T2-W MRI can be useful to improve the radiologist's performance. Nevertheless, multiparametric MRIs such as MRS, DCE and DWI are less available in hospitals than conventional modalities such as T1-W and T2-W. According to Bittencourt *et al.* [57] T2-W imaging is currently the backbone of prostate imaging. This is further confirmed by a recent study conducted by Leake *et al.* [58] where T2-W MRI (together with T1-W MRI) is the most popular imaging modality used as part of general hospitals protocol in current practice of prostate MRI. Moreover, a study by Tempany and Franco [59] states that the T2-W MRI modality is very well established and frequently used analysis compared to the other modalities.

1.5 Computer Aided Detection

In the last 30 years, scientists have been investigating the prospect of integrating computer technology in clinical practices. Although MRIs are very popular for investigation of the internal anatomical body structures, with large amount of images produced every day, analysing each image manually is impractical. Computer Aided Diagnosis/Detection (CAD) is a computerised procedure that can assist radiologists in the interpretation of medical images [60] and has become one of the most popular research subjects in medical imaging. CAD technology has been widely applied in many parts of human body and recently has been successful in breast mammography, colon, brain and lung imaging. The main role of CAD system in radiology practice is assisting radiologists in diagnostic decision making.

Previously CAD systems based on single modality were seen to be a step forward in medical image analysis research. The use of multimodality or multiparametric in CAD systems is becoming popular as more information can be extracted to help radiologists in image interpretation for diagnostic decision making. However, CAD systems using multiparametric MRI produce larger number of images, lack of diagnostic standard procedures, cumbersome for both patient and physician and costly [61]. As a result, radiologists are no longer able to analyse all those images tenaciously and diagnostic results will suffer from observers variability (e.g. results may vary depending on radiologist's expertise and experience). Therefore, the CAD system is needed not

only to eliminate variability among radiologists, but to speed up the analysis of the images and diagnosis decision making.

1.6 The Scope of this Thesis

This study was conducted only within the prostate PZ using a T2-W MRI dataset. Although the development of CAD systems using multiparametric MRI is favored by most researchers, as mentioned early multiparametric MRI is costly, time consuming, and cumbersome for both patient and physician [61]. From a clinical point of view, multiparametric prostate MRI does not have standardised diagnostic criteria. Moreover, a larger number of images data must be analysed in each case and can be computationally expensive and time consuming [45]. In fact, T2-W MRI (and T1-W MRI) is the most popular current practice in prostate imaging [58]. We are aware the fact that using T2-W alone is not sufficient, but that T2-W classification will form a solid basis for a multimodality MRI based system. This study does not attempt to improve the performance of CAD-PC based on multimodality MRI but aims to investigate the performance of CAD-PC using a large number of texture descriptors in T2-W MRI alone within the PZ towards inclusion of a more general multimodality MRI classification platform.

On the other hand, this study was concentrated within the PZ because approximately 75% to 80% of prostate cancers arise within the PZ, and tumours that arise within the this region are more aggressive than those arising in the CZ [12, 15, 16, 18]. Therefore, this project investigated a CAD for prostate cancer detection for the PZ in T2-W MRI.

1.7 Thesis Aims and Objectives

The overall aim of this thesis is to develop CAD systems for prostate cancer detection using single modality T2-W MRI within the PZ. For this purpose segmentation and classification techniques based on texture analysis were thoroughly investigated. For texture analysis, it is important to identify features which can discriminate the characteristics of benign and malignant regions within the prostate's PZ. On the other hand, classification models are built to learn the statistical appearances for both regions in

order to make sure that the developed CAD systems are able to make an accurate prediction/classification. In order to achieve these, the project is divided into a set of objectives:

- To develop a 2D model that can automatically estimate the boundary of the PZ and CZ. Since segmenting the PZ manually is time consuming, it is necessary to develop a method which can automatically segment the prostate PZ by modeling a *priori* general knowledge of radiologists.
- To investigate the texture appearance of malignant and benign regions within the PZ, both from clinical and statistical point of view.
- To identify robust texture descriptors to differentiate malignant and benign regions within the prostate's PZ. There are at least 42 CADs for prostate cancer detection which have been developed in the last 15 years according to the recent survey conducted by Lemaître *et al.* [62] in 2015. Nevertheless, none of those studies have tried to identify the most discriminant texture features in T2-W MRI.
- To investigate the appropriate window sizes (or kernel sizes) in CAD for prostate cancer development. One of the important considerations in feature extraction is selecting the most appropriate window sizes which can change the feature value within a corresponding region. In this project 9 different window sizes (3×3 to 19×19) were investigated to answer this question.
- To develop unsupervised and supervised CAD methods for prostate cancer detection within the PZ in T2-W MRI. Many Studies [46, 63, 64] have reported the limitation of CAD systems using single T2-weighted (T2-W) MRI including weak texture descriptors and noise. In fact, Tiwari *et al.* [65] suggests that T2-W MRI texture features alone might not be sufficient to identify prostate malignancy. Therefore, the use of multiparametric MRI in developing CAD systems is favored by most researchers searching to improve the performances of their methods. However, this thesis investigates CAD based on T2-W MRI due to the issues of multiparametric MRI discussed before.
- To evaluate and compare the CAD methods developed in this study with the existing methods in the literature. To qualitatively assess the reliability of the proposed methods in this project. The results were compared with the existing CADs based on mono and multiparametric MRI.

1.8 Thesis Outline

The remaining chapters of this thesis are organised as follows:

- In Chapter 2, a literature review of texture analysis in biomedical imaging will cover statistical, structural, signal processing and model based methods. In addition, a set of texture descriptors which are commonly used in texture analysis will be reviewed.
- Chapter 3 summarises some of the existing CAD methods based on mono and multiparametric MRI in the literature. This covers the overview of the work flow for each of the components in CAD as well as texture features and classifiers employed.
- In Chapter 4, an unsupervised CAD method using a small number of texture features to segment malignant regions within the PZ is proposed and evaluated.
- Chapter 5 describes a supervised CAD method using 215 texture features. The performance of each texture will be evaluated based different window sizes. From the evaluation results, this chapter will reveal the most discriminant features and the effects on the proposed method of using different window sizes and classifier models.
- Chapter 6 investigates how textons can be used to model the statistical appearance of benign and malignant regions, and proposes a CAD method using textons. This chapter demonstrates the variation of the results when using different window sizes, different number of textons used to represent benign and malignant classes and different classifiers used to build predictive models.
- Chapter 7 will compare the performances of the proposed methods with the existing methods in the literature. This enables us to qualitatively assess whether the proposed methods achieved similar results to the state-of-the-art or not.
- We summarise Chapter 1 until 7 covering discussions and study limitations in Chapter 8.
- Conclusions and future work of this thesis will be covered in chapter 9 including novel contributions and a list of publications arising from this thesis.

1.9 Summary

In this chapter we have briefly presented the fact and figures of prostate cancer covering the world, Europe and the UK as well as the aim, objectives and scopes of this thesis. From clinical point of view:

- The causes of prostate cancer remains uncertain.
- The number of prostate cancer cases is expected to increase globally.
- The incidence in more developed countries is higher than in less developed countries due to better diagnosis practices, hence early detection and better treatment plans can be provided to the patients.
- TRUS guided prostate biopsy is currently the most accurate screening method with accuracy range between 70% to 80% whereas DRE and PSA tests are associated with low sensitivity and overdiagnosis, respectively.
- MR imaging is becoming popular and often being integrated with some of the clinical practices. Nevertheless it is used to identify prostate cancer staging, rather than to detect or diagnose the presence of malignancy.

On the other hand, from computer vision point of view:

- This thesis aims to develop CAD methods for prostate cancer detection using image analysis and machine learning techniques.
- There is a need of CAD systems for prostate cancer detection or diagnosis to assist radiologists or clinicians in making diagnostic decision.
- The development CAD in prostate cancer is an active research field and consistently received attentions from the medical image analysis community.

Chapter 2

Texture Analysis

Texture analysis is one of the most important topics in computer vision due to its wide range of applications, such as biomedical image analysis, industrial inspection, satellite image analysis, etc. Due to the degree of complexity and diversity of textures, it is impossible to provide a universal definition of textures [66]. Despite this lack of a universally agreed definition, scientists agree that one of the key purposes of texture analysis is for pattern recognition. The first section of this chapter will briefly review some definitions of texture from different perspectives. Subsequently, this chapter elaborates four main problem domains in texture analysis and four texture analysis techniques motivated from statistical, geometrical, signal processing, model-based and physiological view points. The last section of this chapter will briefly review some texture descriptors for texture representations.

2.1 Texture Definitions

Texture is one of the most popular properties to characterise objects or regions in an image. Although it is difficult to give a universal definition, in the last 30 years many authors have attempted to define textures from different perspectives. In general, the texture refers to surface characteristics and appearance of an object in terms of size, shape, density, arrangement and proportions of its elementary parts [67]. In computer vision, Shapiro and Stockman [68] define a texture as a set of metrics which contain information about the spatial arrangement of color or intensities in an image or selected region of an image. This means, a texture usually has statistical properties, similarity in structures and degree of randomness [68, 69]. From a geometrical point of

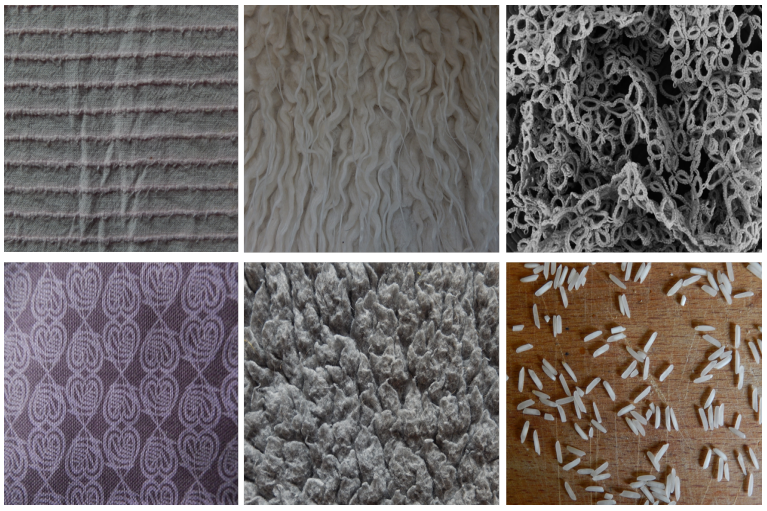

FIGURE 2.1: Examples of three different textures: from left column to right column homogenous, weakly-homogeneous and inhomogeneous, respectively.

view a texture is defined as a set of primitive texels in some regular or repeated pattern [70], where in statistics a texture is the arrangement of intensities in a region which can be quantitatively measured [71]. On the other hand, Taylor *et al.* [72] defined texture as a self-similarity pattern which is represented by similar signal statistics, whereas Cross and Jain [70] suggested that texture is a stochastic, possibly periodic and two-dimensional image field. Finally according to Levine [73] texture can be seen in terms of uniformity, density, roughness, regularity, linearity, frequency, phase, directionality, coarseness, randomness, fineness, smoothness and granulation.

Figure 2.1 shows examples of three different textures. Textures can be divided into three categories according to their properties [72], namely homogenous, weakly homogeneous and inhomogeneous. Homogenous texture has visible repetitive structures and weakly-homogeneous texture involves local spatial variation in textural elements which leads to less repetitiveness. On the other hand, inhomogeneous texture has no repetitive structures and self similarity is absent or random.

2.2 Problem Domains in Texture Analysis

According to Ojala *et al.* [74, 75], there are four main problem domains in texture analysis: (a) texture classification, (b) texture segmentation, (c) texture synthesis and (d) shape analysis.

2.2.1 Texture Classification

Texture classification consists of two stages: learning phase and recognition phase. In the learning phase, the aim is to build a model for the content of each texture class present in the training data. The texture contents in the training data are a set of texture descriptors extracted using a chosen feature extraction method. These texture descriptors can characterise the textural properties of an image, such as similarity, contrast, roughness, homogeneity, orientation, etc. Once the model of the texture content for each texture class was built, in the recognition phase, the texture content of the unknown sample is first described with the same texture analysis method in the learning phase. Subsequently, the textural features of the unknown sample are compared to those of the training images with a classification algorithm, and the sample is assigned to the category with the best match [74, 75].

In biomedical image analysis, Sharma *et al.* [76] used texture-primitive features (e.g. first- and second-order statistical features) for texture classification of different tissues in brain, abdomen and liver. A model for each tissue type from each texture class (e.g. bone/skull, abnormal and normal regions) was built before an Artificial Neural Network (ANN) classifier was employed to classify the unknown sample. Similarly, Wu *et al.* [77] used texture classification to differentiate ultrasonic liver images into the three classes (normal, hepatoma, and cirrhosis) whereas [78, 79] performed texture classification in benign and malignant regions within the prostate gland. On the other hand, in a texture volumes classification Reyes Aldasoro and Bhalerao [80] proposed a multiresolution volumetric texture segmentation (M-VTS) method applied on MRI of human knees. The method extracts texture information from the Fourier domain via subband filtering using an orientation pyramid. A novel Bhattacharyya space was used to select the most discriminant features before there were used to build an oct tree. The classification process was performed at a chosen level of a lower spatial resolution. Subsequently, all labels of the classified voxels are then projected to lower levels of the tree where a boundary refinement procedure is performed.

2.2.2 Texture Segmentation

The basic idea of texture segmentation is to group or cluster pixels based on their properties such as contrast, homogeneity, etc. Pixels that represent the same texture

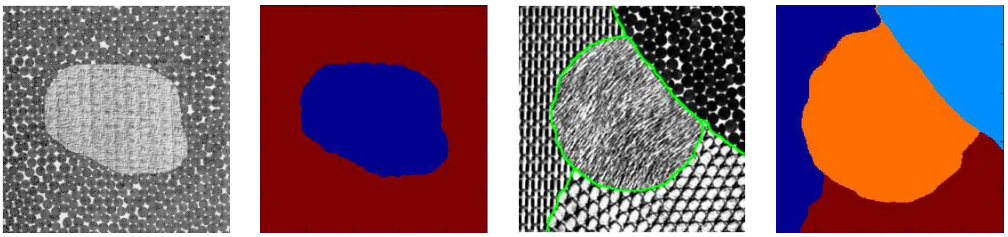

FIGURE 2.2: Example of textures taken from Brodatz album [82] and their segmentation results.

should be grouped together so that further analysis can be done on the various texture regions [81]. However, this is not always the case because there are many cases where two or more different textures have very similar properties leading to incorrect segmentation results. Another challenge of texture segmentation is to identify the boundaries between regions when the regions have similar textures. Texture segmentation can also be quite difficult when there is noise affecting the differentiation between textures. Figure 2.1 shows an example of texture segmentation taken from the Brodatz album [82].
According to [83] there are six most commonly used segmentation techniques in CAD systems:

1. Contour and shape based (e.g. Active Contour, Level Set, Graph Searching, Atlas-based, Deformable models).
2. Machine learning based (e.g. Support Vector Machine (SVM), k-Nearest Neighbors (k-NN), Fuzzy C-Mean (FCM), K-means).
3. Region based (e.g. Thresholding, Edge-based, watershed, split and merge).
4. Statistical based (e.g. Markov Random Field (MRF) and Gaussian Mixture Model (GMM)) .
5. Multiresolution based analysis (e.g. Discrete Wavelet Transform (DWT)).
6. Hybrid and soft computing methods (e.g. Level Set + Artificial neural network (ANN), Fuzzy C-means + DWT).

The most popular segmentation techniques in biomedical imaging fall under the supervised and unsupervised machine learning based techniques, and contour and shape based methods. The level set technique has been applied to several human organs (e.g.

brain, cardiac, prostate, breast, etc). Dubey *et al.* [84] proposed a semi-automatic segmentation method for MRI brain tumors. Firstly, the method generated a tumor probability map by classifying each voxel into the tumor or background class using intensity-based fuzzy c-means. Subsequently, the tumor probability map was used to locally guide the propagation direction of the level set.

Recently, Yeo *et al.* [85], proposed a level-set segmentation method using active contour modelling applied to synthetic and real images (e.g. brain and knee MRI and carotid CT image). The proposed method consisted of an image attraction force, which was used to propagate contours toward object boundaries, and a global shape force, which deforms the model according to the shape distribution learned from a training set. On the other hand, Sachdeva *et al.* [86] proposed a method which used intensity and texture information (extracted from Gray-Level Co-occurrence Matrices) present within the active contour to overcome weak or diffused edges in an image. for Finally, graph searching techniques have been studied for the segmentation of biomedical images of brains and knees [87, 88].

2.2.3 Texture Synthesis

In 1984, Lewis [89] introduced the idea of texture synthesis for digital painting. A couple of years later, Gagalowicz and Ma [90] proposed a method to synthesise 3D textures. Later, texture synthesis gained popularity amongst researchers and several authors [91–93] have successfully developed robust texture synthesis algorithms. According to [91, 94] texture synthesis is an alternative way to create textures and a process of computationally constructing a large digital image from a small digital sample image by considering the properties of its structural content. The output of a synthesised image aims to have the following criteria: the synthesised texture should be as similar as possible to the sample, should not have visible artifacts and no structure repetition of the sample in the synthesised texture [94]. Potential applications of a successful texture synthesis algorithm are broad, including occlusion fill-in, lossy image and video compression and foreground removal [93]. Some popular texture synthesis methods are pixel-based [91, 95] and patch-based [96].

In medical image analysis, a successful texture synthesis algorithm can provide a better anatomical image quality as noise is absent in the synthetic image. Bochud *et al.* [97] proposed a method to generate medical texture synthesis methods in mammography derived from the lumpy background technique described by Rolland and

Barrett [98]. In a similar application, Taylor [99] proposed a statistical-based model to synthesise textures by modelling the texture of the sample parametrically using k-means clustering algorithm. Subsequently, a seed image was formed and a sliding window used to create synthetic texture based on the texture model built previously.

2.2.4 Shape Analysis

Shape is another important property to characterise textures. The main goal of shape analysis is to learn the structures and contours of the texture content in an image. Shape analysis has two main steps: shape correspondence (computes correspondent locations between geometric shapes) and statistical analysis (measurements of structural change at correspondent locations) [100]. According to [101], shape analysis considers the area, circularity, eccentricity, orientations and a set of algebraic moment invariants of the texture. In addition, Mingqiang *et al.* [102] suggested that shape analysis measures the following properties: rectangularity, solidity, symmetricity, profiles, hole area ratio and center of gravity. The representation of shapes which cover the shape feature extraction techniques can be referred in [102].

In medical image analysis, shape analysis is used to study the geometrical properties of structures obtained from different imaging modalities [100]. Shape analysis is able to precisely locate morphological changes between different populations (e.g. female vs male and young vs elderly). Shape analysis is used in many different studies such as faces, brain structures and mammography calcification. In a study of facial feature detection, Chen and Tiddeman [103] used the principal direction of the spatial shape for the estimation of face direction. In brain structures, Martin *et al.* [104] used shape features to analyse the shape deformation of structures in the human brain by modelling the brain's elastic properties using physical and experimental modes. Finally, in mammography calcification, Shen *et al.* [105] developed a set of shape factors to measure the roughness of contours of calcification to classify them as malignant or benign.

2.3 Texture Analysis Techniques

In texture analysis, many different approaches have been developed and it is impossible to review all of them thoroughly. Nevertheless, this section will briefly review some of

the commonly used methods for extracting texture features. They will be divided into four categories: statistical methods, structural methods, signal processing methods, and model-based methods.

2.3.1 Statistical Methods

The spatial distribution and intensity of pixel values in an image are among the most important characteristics in describing texture. Feature extraction based on statistical analysis is one of the earliest methods developed to capture the content of the texture in an image. According to Haralick [106], a texture contains two basic dimensions: the first dimension describes the primitives (or texture elements, texels) out of which the image texture is composed, and the second dimension is for the description of the spatial interaction between the primitives of an image texture. This means, the first dimension contains information about the local properties and the second dimension contains the spatial information. Statistical methods try to characterise a texture by utilising these properties such as intensity and spatial information. The first-order statistical features are among the easiest properties that can be extracted from textures; for examples mean, median, variance, skewness, kurtosis, etc. These features can be directly computed from an image by taking the intensity or grey level values within a defined window size and provide information related to the intensity or grey level distribution of the image.

However, the first-order statistical features do not provide information about the spatial information among the grey levels, such as whether all low-value grey levels are positioned together or are interchanged with the high-value grey levels [107]. The second-order statistical features take the spatial information into account. Haralick *et al.* [108] developed grey level co-occurrence matrix (GLCM) to describe two-dimensional spatial relationship of grey levels within a given distance and orientation [109]. The GLCM is defined as the joint probability of occurrence of two grey level values at a given offset both in terms of distance and orientation [110, 111]. This means, each value of the co-occurrence matrix element presents relative frequency with two adjacent grey levels. From this matrix, a set of texture descriptors can be extracted such as homogeneity, contrast, energy, correlation, etc. In addition, Haralick *et al.* [108] formulated other features such as inverse difference and inverse difference moment normalized. Conners *et al.* [112] recommended two additional features in GLCM namely cluster shade and cluster prominence.

There are a few variations of GLCM such as grey level difference matrix (GLDM), grey level run length matrix (GLRLM) and spatial gray-level dependence matrix (SGLDM). GLDM is based on the occurrence of pixel pairs given an absolute difference in gray level and separated by a specific distance [113, 114]. Five textural features are measured using GLDM: contrast, angular second moment, entropy, mean, and inverse difference moment. On the other hand, the GLRLM [115] is based on computing the number of gray-level runs of several lengths. A grey level run is a set of linearly adjacent pixel points having the same grey level value. The length of the run is the number of pixel points in the run. A set of textural features are measured from the matrix, such as short-run emphasis, long-runs emphasis, gray-level nonuniformity, run-length nonuniformity, and run percentage [115]. Finally the SGLDM [113, 114] is based on an estimation of the second-order joint conditional probability density functions (PDF) for a given orientation. The PDF function is the probability that two pixels located with a given distance and direction, have corresponding gray levels. This means, SGLDM aims to construct co-occurrence matrices to reflect the spatial distribution of gray levels in the region of interest.

2.3.2 Structural Methods

Structural based methods (also known as geometrical methods) attempt to analyse the structural properties of the primitives in order to characterise the texture [106]. In structural methods, texture is viewed as a set of many textels, or also known as textural elements, arranged according to some placement (e.g. orientations and distance) rules [116]. In texture analysis, many structural based methods are under the problem domain of shape analysis because these methods tend to use structural properties such as area, perimeter, eccentricity, orientation, elongation, magnitude, compactness, etc [117, 118].

Goyal *et al.* [119] proposed a robust method to calculate compactness by improving perimeter feature to be invariant to shift and rotation. By making the perimeter invariant to shift and rotation, it helps compactness invariant to translation, rotation and scaling. Ahuja [120] proposed a model for defining the neighbourhood of a point called Voronoi tessellation. The region enclosed by a point's Voronoi polygon was assumed to possess intuitively appealing characteristics and was regarded as the neighbourhood of the point [109]. On the basis of that, Tuceryan and Jain [121] developed a texture segmentation algorithm based on Voronoi tessellation (consists of

many Voronoi polygons) to model the texture. Subsequently, the method computes shape features from Voronoi polygons and used these features to identify the interior and the border regions of the textures.

In another study, Matsuyama *et al.* [122] detect texture periodicity by finding the peaks from the Fourier spectrum of a textured image. Similarly, Liu *et al.* [123] detect the peaks but from the output of autocorrelation functions of a texture. The structural methods of Serra [124] and Lu and Fu [125] developed a tree grammar syntactic approach for a texture representation. Firstly, a texture is divided into a small 9×9 square windows and windows belonging to the same texture pattern are then characterized by a tree grammar. The constructed tree grammar was used for synthesis as well as discrimination. The advantage of the structural approach is that it provides a good symbolic description of the image; however, this feature is more useful for synthesis than analysis tasks [126].

2.3.3 Signal Processing Methods

According to a psychological study conducted by Bruce *et al.* [127], there is a strong relationship between visual perception and spatial frequency in human brain. Stromeyer and Klein [128] suggested that the human visual system is sensitive to textures due to the existence of channels which are selectively sensitive to different ranges of spatial frequencies. Based on this proposition, many authors have developed signal processing methods for texture analysis. Many of these methods utilise filters to capture frequency information of the texture to extract its properties. Since the appearance of textures can be very complex, frequency content of the image also can be very complex too, therefore designing robust filters to be able to capture these complex frequencies can be very difficult.

Laws [129] designed a set of 2D masks derived from three simple 1D filters for flats, edges and spots detection. Each of these 1D masks will be convolved to provide a set of symmetric and anti-symmetric centre-weighted masks with all but the level filters being zero sum. The developed filters are very successful in flats, edges, spots, and ripple detection. Furthermore, Law [129] introduced another class of spatial filters called moments, which correspond to filtering the image with a set of spatial masks used to extract texture features. Moment-based features were successfully applied in texture segmentation by Tuceryan [130]. Another way to capture frequency information is via linear transformation of the local neighborhood vector proposed by

Unser and Eden [131]. For this method, a bank of local convolution masks are stacked to form a transformation matrix, then each row of the transformed matrix represents a mask that will be used to filter the input texture [131]. In contrast, Rosenfeld and Thurston [132] introduced masks for edge detection based on edge density on the basis that fine textures tend to have a higher density of edges than coarse textures [117].

Other filters based on signal processing methods are Fourier [133], Gabor [134, 134] and wavelet transforms [135, 136]. These methods represent an image in a coordinate system space which has an interpretation similarity with the characteristics of a texture. The Fourier transform captures the global frequency content of an image without any references to localisation in the spatial domain where Gabor filters provides references for spatial localisation but one of the limitations is that the filters could not localise a spatial structure in natural textures [74, 75, 117]. In multiresolution analysis, wavelet transforms solved this issue using a window function where the window size changes as the frequency changes [137] which is similar to difference-of-Gaussians [138], 2D Gabor filters [134, 134] and filter banks [139]. These methods filter an input image with a bank of filters and each filter has a specific frequency and orientation. The resulting filtered images are the texture features representing different frequencies and orientations of the texture. As the number of filters increases, the dimensionality of the data is also increasing; in some texture segmentation methods dimensionality can be reduced using Principal Component Analysis (PCA), feature selection and only considering filtered images which have high energy.

2.3.4 Model Based Methods

Model based texture analysis methods try to build a model that can represent the texture of an image via several techniques such as generative image model and stochastic model [140]. Several authors mentioned that one of the problems with most stochastic models is estimating their parameters, which could be computationally expensive [141, 142]. One of the most popular model based methods is Markov random fields (MRF) [143]. The MRF model assumes that each pixel's intensity within a region of interest in the image is highly dependent on the intensity values of its local neighbourhood [109]. This means the MRF model represents a set of connected nodes (or pixels) in a graph. Mandelbrot [144] proposed a method to model textures of an image with fractals using fractal dimension function. The function can characterise

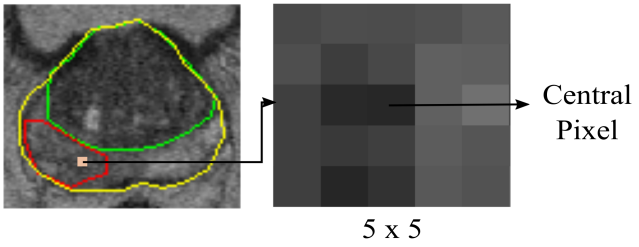

FIGURE 2.3: Example of 5×5 window used to calculate the features.

self-similarity and roughness of image texture [145]. In contrast to MRF models, fractals have high power in low frequencies and fractal dimension is scale invariant. Finally Jojic *et al.* [146] proposed a novel appearance and shape models called *epitomes*. The *epitome* is a smaller image containing a miniature and condensed version of the original image such as texture and shape properties [109].

2.4 Texture Descriptors for Texture Representation

Texture based features are powerful for texture representations and have shown to be effective for texture classification and segmentation. By analysing an image texture, we can capture its appearance, orientations, scales, distributions, etc. This section will review some of the most popular texture features mainly used in CAD for prostate cancer studies.

2.4.1 First-Order Statistical Features

Generally, first order-statistical features compute local features by analysing the distribution of gray values at each point in the image. Features under this category are the basic texture descriptors used in texture analysis and estimate properties (e.g. average and variance) of an individual pixel by replacing the value of the central pixel with the result of its neighbouring values within a given window size. All features under this category do not capture the spatial information between grey levels. As a result, textures with similar grey level distributions but different spatial arrangements could not be distinguished. Figure 2.3 shows an example of 5×5 window used to

compute each of the features described below. Let $A_{(i,j)}$ denotes each element (or pixel) within a $m \times n$ sliding window (W), where m and n are row and column, respectively. Each element within the sliding window has (i, j) position. W contains $m \times n$ number of elements (N). The resulting value will be replaced into the central pixel. The following equations are the list of first-order statistical features used in this study.

- Mean is calculated by adding up all the grey level values within the W and dividing that sum by the total number of grey levels.

$$(\mu) = \frac{1}{N} \sum_{i=1,j=1}^{N} A_{(i,j)} \tag{2.1}$$

- Median is the middle value in a list of ascending order grey level values (X) within the W.

$$med \text{ is the } (\frac{N+1}{2})^{th} \text{ element of X} \tag{2.2}$$

- Variance indicates how widely the grey level values within the W vary.

$$var = \frac{1}{N-1} \sum_{i=1,j=1}^{N} A_{(i,j)} - \mu^2 \tag{2.3}$$

- Standard Deviation is the square root of the variance.

$$std = \sqrt{\frac{1}{N-1} \sum_{i=1,j=1}^{N} A_{(i,j)} - \mu^2} \tag{2.4}$$

- Skewness is a measure of the asymmetry of the probability distribution of the grey level values about its mean within a given window size.

$$sk = \frac{\sum_{i=1,j=1}^{N} \left(A_{(i,j)} - \mu\right)^3}{(N-1)\,\sigma^3} \tag{2.5}$$

- Kurtosis measures whether the data are peaked or flat relative to a normal distribution.

$$kur = \frac{\sum_{i=1,j=1}^{N} (A_{(i,j)} - \mu)^4}{(N-1)\,\sigma^4} \tag{2.6}$$

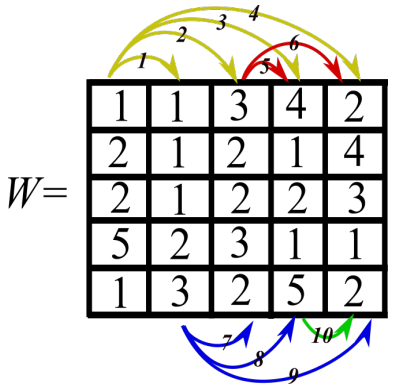

FIGURE 2.4: Pair column correlation in a 5×5 window.

- Mean absolute deviation calculates the average difference between each grey level value and the mean within the W.

$$mad = \frac{1}{N} \sum_{i=1,j=1}^{N} A_{(i,j)} - \mu \tag{2.7}$$

- Median absolute deviation is the middle value in Y. Note that Y is in an ascending order.

$$mead \text{ is the } (\frac{N+1}{2})^{th} \text{ element of } Y \tag{2.8}$$

$$Y = A_{(i,j)} - med \tag{2.9}$$

- Average of rank correlations can be calculated by computing the correlation for each pair of columns then summing up the total correlations before dividing it by the number of paired columns.

$$avgCorr = \frac{1}{N_{pc}} \times \frac{\sum_{i=1}^{N}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{N}(X_i - \bar{X})^2 \sum_{i=1}^{N}(Y_i - \bar{Y})^2}} \tag{2.10}$$

where N_{pc} is the number of pairs of columns which is 10 based on the example in Figure 2.4 and X and Y are the first and second columns, respectively in a of pair column.

- Local contrast represents the difference between the highest and lowest grey level values within the W.

$$lc = max(W) - min(W) \tag{2.11}$$

- Local probability computes the probability of occurrence for each grey level within the W.

$$lp = \frac{\#(A(i,j) = k)}{m \times n} \tag{2.12}$$

where $\#(A(i,j) = k)$ is the number of pixels at the k^{th} grey level in a $m \times n$ image.

- Percentile 25 indicates the grey level distribution within W under a given threshold (e.g. 25%)

$$p25 = \begin{cases} \text{the } idx^{th} \text{ element of } X, & \text{if } idx \text{ is a whole number} \\ \text{the } round(idx)^{th} \text{ element of } X, & \text{otherwise} \end{cases} \tag{2.13}$$

$$idx = N \times 0.25 \tag{2.14}$$

- Percentile 75 indicates the grey level distribution within W under a given threshold (e.g. 75%)

$$p75 = \begin{cases} \text{the } idx^{th} \text{ element of } X, & \text{if } idx \text{ is a whole number} \\ \text{the } round(idx)^{th} \text{ element of } X, & \text{otherwise} \end{cases} \tag{2.15}$$

$$idx = N \times 0.75 \tag{2.16}$$

2.4.2 Second-Order Statistical Features

One of the most popular second-order statistical features is a set of features introduced by Haralick *et al.* [108]. In comparison to the previous ones, these features consider the spatial relationship between grey levels within a given distance (d) and orientation (θ). By definition, GLCM is a tabulation of how often different combinations of grey levels

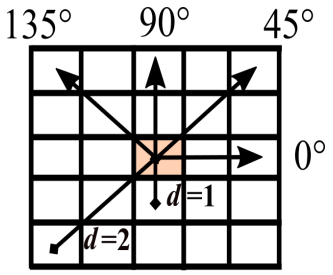

FIGURE 2.5: An illustration of orientation (θ) and distance (d).

occur in an image. This means, each value in the GLCM represents the probability occurrence of a pair of grey level values at a given distance (d) and orientation (θ). In many texture analysis methods, combining first and second-order statistical features are popular because they describe the grey level distribution and spatial relationship. Figure 2.5 illustrates the spatial relationships of pixels that are defined by this array of orientations and the distance from the pixel of interest (the central pixel). Note that, d can be calculated by counting the number of pixels adjacent (excluding the central pixel) from the central pixel. For example, $d = 1$ means there is one adjacent pixel and $d = 2$ involves two adjacent pixels. Second-order statistical features involves two stages: the first is to construct the GLCM and the second is to compute features from the constructed GLCM. To create the GLCM, the following steps are required:

- Create a framework matrix based on the grey level values in the input image
- Choose d and θ (e.g. $d = 1$ and $\theta = 0°$)
- Count the occurrences and fill in the framework matrix to create the first co-occurrences matrix (M)
- Create a transpose matrix (M^t)
- Add the corresponding values in M and M^t to create a symmetrical co-occurrences matrix (M^s)
- Normalize the M^s to turn it into probabilities. This can be done by dividing each element in the M^s by the number of total occurrences in M^s. Here the GLCM (G) is constructed.

Figure 2.6 shows an example of how to construct M from an input image using $d = 1$ and $\theta = 0°$. The frequency value of M at the position $(1, 1)$ is 2, hence $M_{(1,1)} = 2$

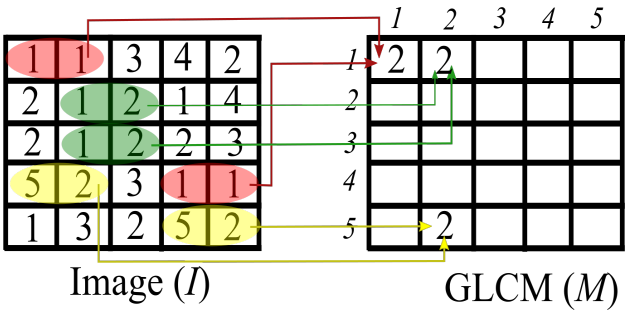

FIGURE 2.6: An example to construct M from I.

because the grey level pair $(1, 1)$ occur twice in I. Similarly, $M_{(5,2)} = 2$ because the grey level pair $(5, 2)$ occur twice in the input image. On the other hand, $M_{(1,2)} = 1$ because the grey level pair $(1, 2)$ occur only once in I. Once M is created, M^t can be created by doing the same process in Figure 2.6 but considering grey level pairs from the opposite direction. For example, M is created by pairing grey levels from left to right and M^t can be created by pairing grey levels from right to the left.

Figure 2.7 shows an example of creating M^t. Once M and M^t are created, a symmetrical co-occurrences matrix (M^s) can be constructed by adding up the corresponding values in both matrix as shown in 2.8. Subsequently, the co-occurrence probability for each value in M^s is calculated to create G. This can be done by dividing each value in M^s by the sum of values in M^s, which means each value in the G represents the co-occurrence probability of a pair of grey levels based on the parameters d and θ. Using the G, Haralick *et al.* [108] proposed the following features for describing texture in an image.

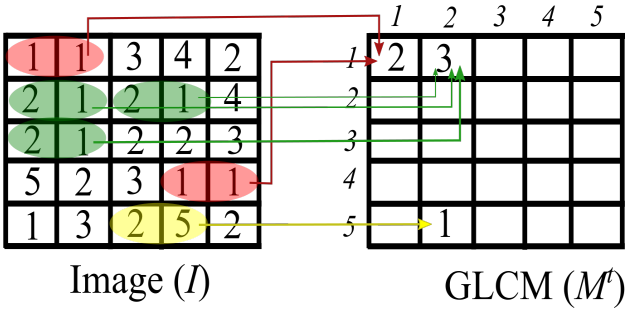

FIGURE 2.7: An example to construct M^t from an input image (I).

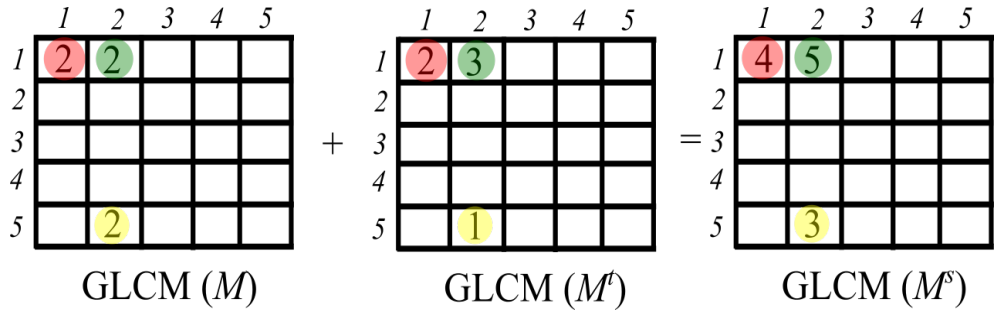

FIGURE 2.8: An example to construct M^s from an M and M^t.

The following notations are used.

- N_g is the number of distinct grey levels in G.
- x and y are the coordinates of the row and column, respectively.
- $p_{(i,j)}$ is the probability value in G at position (i, j).
- avg is the mean of $p_{(i,j)}$.
- μ_x,μ_y,σ_x,σ_y are the means and standard deviations of p_x and p_y. The value of $p_x(i)$ can be obtained by summing up all the values at the i^{th} row. Similarly, the value of $p_y(j)$ can be calculated by adding up all the values at the j^{th} columns.

Therefore:

$$p_x(i) = \sum_{i=0}^{N_g-1} p_{(i,j)} \tag{2.17}$$

$$p_y(j) = \sum_{j=0}^{N_g-1} p_{(i,j)} \tag{2.18}$$

$$\mu_x = \sum_{i=0}^{N_g-1} i \times p_x(i) \tag{2.19}$$

$$\mu_y = \sum_{j=0}^{N_g-1} j \times p_y(j) \tag{2.20}$$

$$\sigma_x = sqrt \sum_{i=0}^{N_g-1} (p_x(i) - \mu_x(i)) \tag{2.21}$$

$$\sigma_y = sqrt \sum_{j=0}^{N_g-1} (p_y(j) - \mu_y(j)) \tag{2.22}$$

$$p_{x+y}(k) = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} p_(i,j) \tag{2.23}$$

where $i + j = k$ and $k = 0, 1, 2..2(N_g - 1)$

$$p_{x-y}(k) = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} p_(i,j) \tag{2.24}$$

where $i - j = k$ and $k = 0, 1, 2..N_g - 1$

Based on these notations, the following features can be extracted from G.

$$\text{Contrast } (con) = \sum_{n=0}^{N_g-1} n^2 \Bigg\{ \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} p_{(i,j)} \Bigg\}, i - j = n \tag{2.25}$$

$$\text{Correlation } (cor) = \frac{\sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} \{i \times j\} \times p_{(i,j)} - \{\mu_x \times \mu_y\}}{\sigma_x \times \sigma_y} \tag{2.26}$$

$$\text{Entropy } (ent) = - \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} p_{(i,j)} \times log(p_{(i,j)}) \tag{2.27}$$

$$\text{Inverse different moment (homogeneity) } (idm) = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} \frac{1}{1 + (i-j)^2} \times p_{(i,j)} \tag{2.28}$$

$$\text{Energy } (eng) = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} (p_{(i,j)})^2 \tag{2.29}$$

$$\text{Sums of squares: variance } (svar) = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} p_{(i,j)} (i - avg)^2 \tag{2.30}$$

$$\text{Sum average } (savg) = \sum_{i=0}^{2N_g-1} i \times p_{x+y}(i) \tag{2.31}$$

$$\text{Sum entropy} \quad (sent) = -\sum_{i=0}^{2N_g-1} p_{x+y}(i) \times log(p_{x+y}(i)) \tag{2.32}$$

$$\text{Sum variance} \quad (svar) = \sum_{i=0}^{2N_g-1} (i - savg)^2 \times p_{x+y}(i) \tag{2.33}$$

$$\text{Difference entropy} \quad (dent) = -\sum_{i=0}^{N_g-1} p_{x-y}(i) \times log(p_{x-y}(i)) \tag{2.34}$$

$$\text{Information measure of correlation 1} \quad (imc1) = \frac{HXY - HXY_1}{max\{HX, HY\}} \tag{2.35}$$

$$\text{Information measure of correlation 2} \quad (imc2) = \sqrt{1 - exp[-2(HXY2 - HXY)]} \tag{2.36}$$

where HX and HY are entropies of p_x and p_y and

$$HXY = -\sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} p_{(i,j)} log(p_{(i,j)}) \tag{2.37}$$

$$HXY1 = -\sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} p_{(i,j)} log(p_x(i) p_y(j)) \tag{2.38}$$

$$HXY2 = -\sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} p_x(i) p_y(j) log(p_x(i) p_x(j))$$

In addition, Haralick [108] further introduced the following features:

$$\text{Autocorrelation} \quad (acor) = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} (i \times j) \times p_{(i,j)} \tag{2.40}$$

$$\text{Maximum probability} \quad (mprob) = \text{MAX } p_{(i,j)} \text{ within } G \tag{2.41}$$

$$\text{Dissimilarity} \quad (diss) = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} i - j \times p_{(i,j)} \tag{2.42}$$

On the other hand, Conners *et al.* [112] suggested cluster shade ($csha$) and cluster prominence ($cprom$) which measure the skewness of the matrix using the following equations.

$$csha = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} (i + j - \mu_x - \mu_y)^3 \times p_{(i,j)} \tag{2.43}$$

$$cprom = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} (i + j - \mu_x - \mu_y)^4 \times p_{(i,j)} \tag{2.44}$$

2.4.3 Local Binary Pattern

Local binary pattern (LBP) is another popular texture descriptor in texture analysis, originally introduced by Ojala *et al.* [74]. In comparison to the first and second-order statistical features, the LBP is invariant to monotonic grey level changes, low computational complexity and convenient multi-scale extension [147]. By definition, the LBP operator is an operator which describes the surrounding of the central pixel by generating a bit-code between the central pixel with its neighbouring pixels [148]. On the other hand, the LBP value is a decimal number which represents the LBP code (binary code) constructed by the LBP operator. The LBP code can be computed based on the following condition

$$neigh_{(i,j)} \begin{cases} 1, \, c_{pxl} \geq \ neigh_{(i,j)} \\ 0, \, c_{pxl} < \ neigh_{(i,j)} \end{cases} \tag{2.45}$$

where $neigh_{(i,j)}$ is the neighbouring value of the central pixel (c_{pxl}) at position (i, j) in I. The LBP code can be constructed by comparing the central pixel with each of its neighbouring pixels. Those values more than the central pixel's value will be set to 1 and 0 to those values less than the central pixel. Figure 2.9 shows an overview of LBP and it summarises the following steps:

- Step 1: Divide an input image (I) into cells (e.g. 20×20)

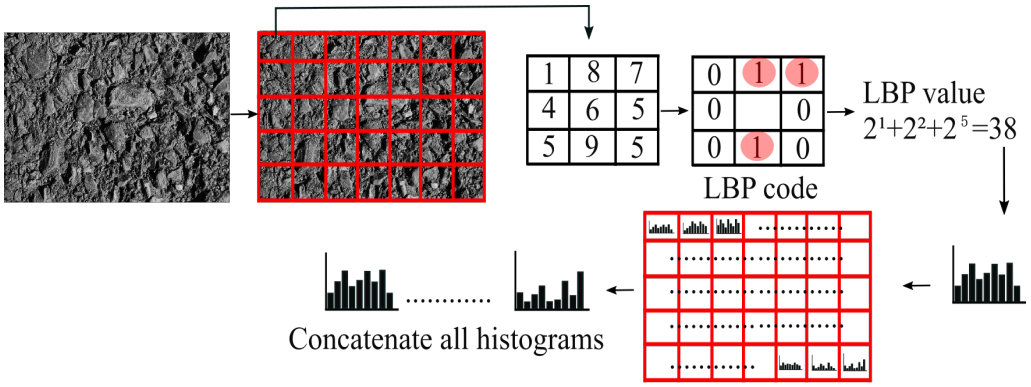

FIGURE 2.9: An overview of extracting LBP features of an image.

- Step 2: Using a 3×3 sliding window (W), compare the c_{pxl} with each of its surrounding pixels (or neighbouring pixels) based on the condition in 2.45 to create its LBP code.
- Step 3: Calculate the LBP value based on clockwise rotation
- Step 4: Repeat step 2 and step 3 until the whole cell is covered (also the whole image). Now the input image is represented in a form of LBP value.
- Step 5: For each cell, compute its histogram by counting the frequency of each distinct LBP code. This means, the histogram of each cell represents the number of occurrences of each LBP value.
- Step 6: Normalise each histogram to sum equal to 1 and concatenate all histograms.
- Step 7: The concatenated histogram will be treated as a feature vector for classification.

Over the last 20 years, LBP has been used in many different applications such as in medical imaging [149], face recognition [150], fingerprint identification [150], texture classification [149], etc. As a result, LBP has been extended to different variations to overcome some of the limitations of the original LBP. For example, although LBP allows a rotation invariant descriptor, it is less sensitive in capturing information about anisotropic structures (e.g. eyes, mouths) in face recognition applications [151]. To capture the anisotropic structures, Liao and Chung [151] proposed elongated local binary pattern (ELBP) which used elongated neighborhood distribution, in contrast to the conventional LBP which uses circular neighborhood. Tan and Triggs [152]

proposed local ternary patterns (LTP) to solve the problem of the sensitivity to noise in uniform textures. Instead of using two threshold values (e.g. 1 and 0), LTP code is constructed based on three threshold values (e.g. 1, 0 and -1). Subsequently, central pixel will be compared to its neighbouring values by including a user-specified threshold. Further, the authors demonstrated more features can be extracted from the LTP code by dividing it into positive and negative parts.

In another study [153], a new technique called median binary patterns (MBP) was proposed for obtaining a noise resistant texture descriptor. In this approach, the median value within a neighborhood is calculated first and compared with the neighbouring pixels including the central pixel to create the MBP code. A similar idea was proposed in [154] and [155]. The conventional LBP was improved to reduce the effect of noise by comparing the neighborhood pixels' intensity values against the local mean pixel intensity, instead of the intensity of the central pixel [149]. On the other hand, Heikkilä *at al.* [156] proposed a reduced dimension LBP called center-symmetric LBP (CSLBP). In this technique, the authors compared only center-symmetric pairs of pixels. Therefore, instead of 2^8 different binary patterns (considering 8 neighbors), this method produces only 2^4 binary patterns (this resulting a shorter histogram). Finally Raja and Gong [157] proposed multiscale LBP to capture multiscale based features which compares the central pixel with its neighbouring values at the number of different scales defined by the user.

2.4.4 Filter Banks

Filter banks are arrangements of low pass, bandpass, and highpass filters used for the spectral decomposition and composition of signals [158]. In texture analysis a filter bank consists of filters with various scales and rotations which aim to capture texture properties at different resolutions and orientations. This means, each filter is designed to capture specific properties of the texture. Subsequently, all information captured by each filter is combined and aims to be a set of robust texture descriptors. In this section, the following bank filters will be discussed: Gabor filters, Leung-Malik filter bank (LM) [159], Schmind filter bank (S) [160] and maximum response filter bank (MRS) [139, 161].

The 1D Gabor filter was introduced by Denis Gabor in 1946 for signal processing [162]. In 1984, Daugman introduced 2D Gabor filters to model the receptive-field

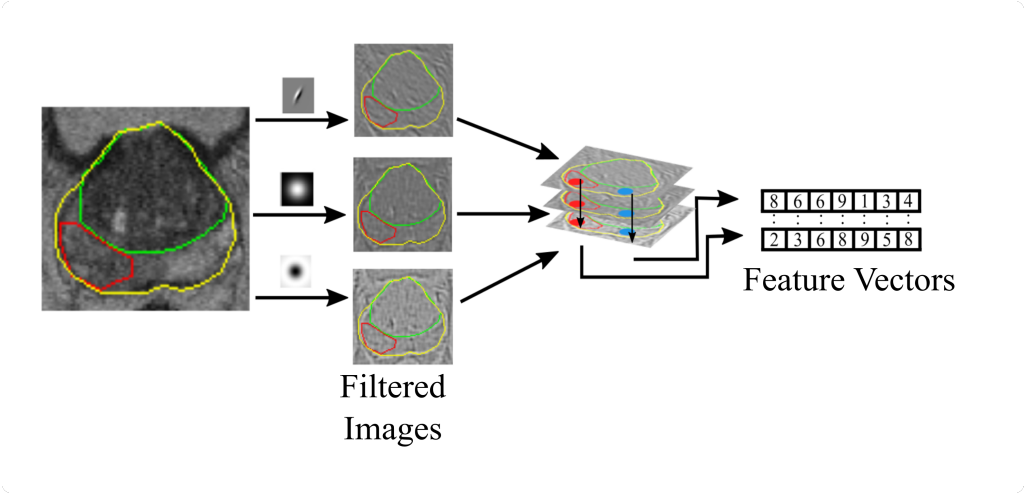

FIGURE 2.10: Feature extraction using filter bank.

profiles of simple cells in the striate cortex [134]. Frequency and orientation representations of the 2D Gabor filters are similar to those of the human visual system which was found to be useful in texture representation. The Gabor filter bank consist of different kernels of different scales (σ), orientations (θ) and wave length (λ). To extract features using Gabor filter bank, an input image is filtered using various kernels via convolution. Each pixel in the filtered image is the response of each filter. Therefore each corresponding pixel of filtered images can be treated as a feature vector dimensioned by the number of filters [163]. Alternatively, a further step can be added by computing statistical features such as mean, energy and standard deviation for each of the filtered images [164]. Subsequently, feature images will be used for texture classification or segmentation. Since this method depends heavily on σ, θ and λ, it can be difficult to find the optimal parameters. Nevertheless, several parameter optimisation approaches of Gabor filter have been proposed in the literature [165–167].

Figure 2.10 shows a typical flow of feature extraction using filter bank. In contrast, the parameters of LM, S and MRS filter banks are already defined by the authors. In terms of feature extraction, these methods share a similar fundamental concept (an input image is filtered by various kernels and the resulting filtered images are treated as image features). The original LM has 48 filters (each has different orientations and scales) consisting of first and second derivatives of Gaussians at 6 orientations and 3 scales (total 36 filters), 8 Laplacian of Gaussian (LoG) filters and 4 Gaussians [159]. On the other hand, Schmid (S) proposed a set of rotationally invariant filter banks which has 13 isotropic 'Gabor-like' filters [160]. The MRS filter banks [168]

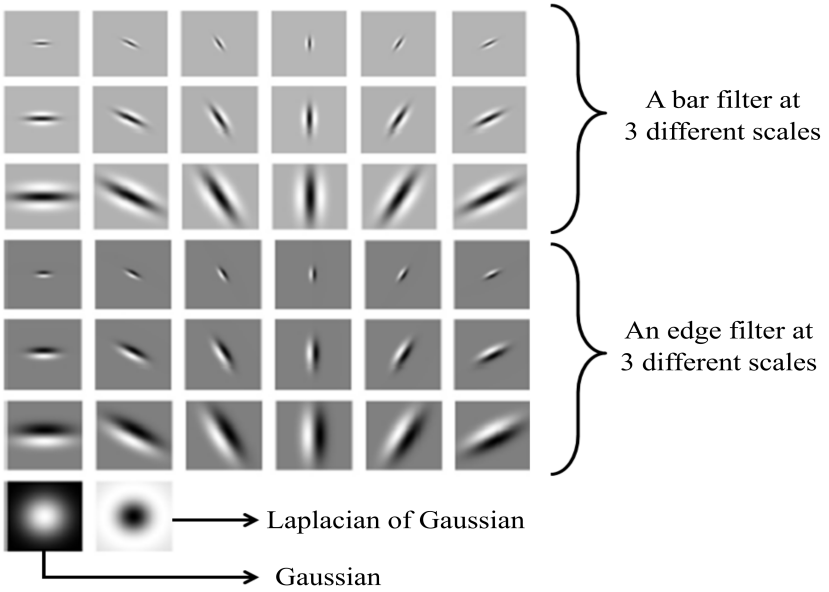

FIGURE 2.11: Filter bank of Varma and Zisserman [1, 168] consists of 38 filters.

consisting of an edge and a bar filter, at 6 orientations ($\theta = 0°, 30°, 60°, 90°, 120°, 150°$) and 3 scales ($(\sigma_x, \sigma_y) = (1,3), (2,6), (4,12)$), Gaussian and LoG filters both with σ =10 pixels (in total 38 responses). To achieve rotational invariance, Varma and Zisserman [1, 168] derive the Maximum Response 8 (MR8) filter bank by taking only the maximum filter response across all orientations together with LoG and Gaussian filters. Therefore, this reduces the number of responses from 38 to 8 (3 scales for 2 filters, plus 2 isotropic). Furthermore, the dimensionality of the filter response space can be reduced by taking the response with the maximum scale and orientation from the maximum responses of edge and bar filter.

Figure 2.11 shows 38 filters of Varma and Zisserman [1, 168] consists of an edge and a bar filters (both at three scales and six orientations), Gaussian and LoG filters. MR8 filter bank can be obtained by taking the maximum response across all orientations for edge and bar filters together with Gaussian and LoG filters.

2.4.5 Textons

Textons refer to fundamental micro-structures in natural images and are considered as the atoms of pre-attentive human visual perception [169]. Unfortunately, 'texton' remains a vague concept in the literature for lack of a good mathematical model

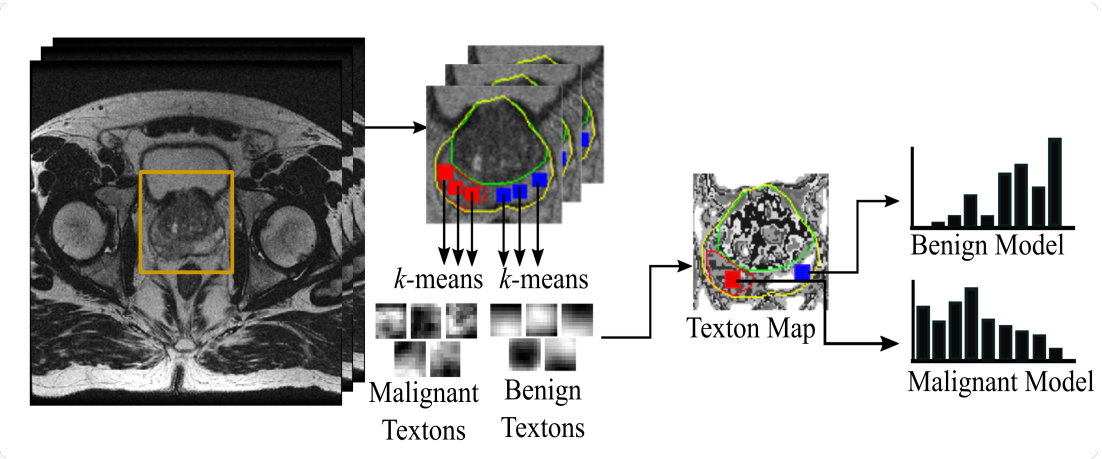

FIGURE 2.12: An overview of generating textons and histogram model for each class.

[170]. In a simple definition, a texton is the cluster centroid generated by the employed clustering algorithm (e.g. k-means). Although the term 'texton' was firstly introduced in the 80's, it did not get much attention until a study of texture classification by Leung and Malik [159] in 2001. Similar studies showing promising results in texture classification conducted by Varma and Zisserman [1, 168] in 2005 and 2009, respectively. On the other hand, in medical image analysis textons have been used in retinal vessel segmentation [171] and lung cancer detection [172]. In many applications, textons are used to model textures in histograms (e.g. different textures can be represented different histograms). Textons can be generated either by clustering the filtered image or by clustering the raw pixels directly from the the original image. In texture classification, results from a study by Varma and Zisserman [161] showed that textons generated from the row pixels of an image produced better results then the ones generated from filtered images. Figure 2.12 summarises the steps to generate textons and histogram models for each class. Note that in this example raw pixels are directly clustered without using filter bank.

- Step 1: Textures in the same class are filtered using filter banks such as Gabor filter bank, LM filter bank, MR filter bank, etc.
- Step 2: All filtered images are aggregated and clustered using k-means clustering algorithm, resulting in a texton dictionary.
- Step 3: Generate a texton map for each of the textures in a class. Resulting a texton map for each texture.

- Step 4: Construct a frequency histogram for each texture from the texton map. Hence, each texture has a histogram which represents the textons distribution within the texture.
- Step 5: The histogram is treated as a feature vector. For example if there are 10 textures in a class, there will be 10 histograms to represent the class.
- Step 6: A classifier is employed to build a predictive model learning the models of each class.
- Step 7: The predictive model is used to test unknown cases.

Alternatively, a study in [1] did not use a classifier to learn the histogram models for each class but employed k-Nearest Neighbor (k-NN) to find the most similar histogram model with the test model in the Euclidean space. Many studies have shown textons [1, 168, 171, 172] are very robust in texture classification. However, one of the main disadvantages of this approach is finding the optimal number of textons which can represent the characteristic of the texture and an optimal patch size needs to be employed.

2.4.6 Gradient Based Features

There are many operators which can be used to extract these features such as Sobel filter, Kirsch filter, etc. These features are popular because they perform well in characterising micro-textures as well as providing more consistent behavior as a descriptor compared to co-occurrence matrices [173]. Figure 2.13 illustrates the components in I

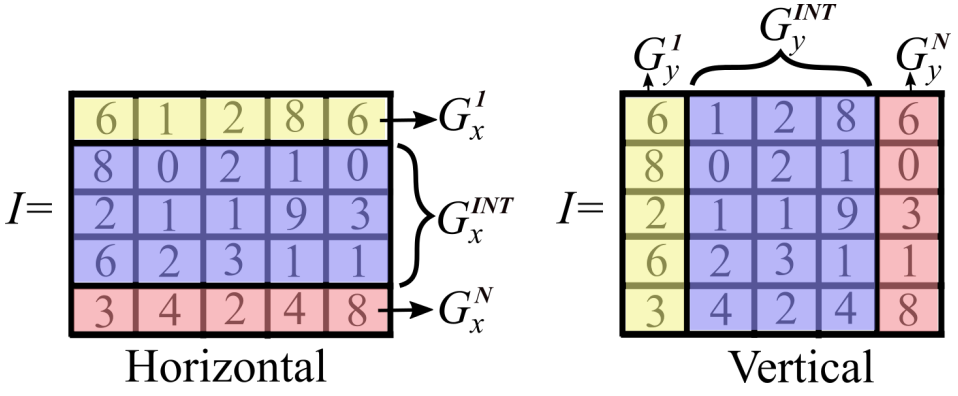

FIGURE 2.13: Components in I based on the calculation of numerical image gradient.

based on the calculation numerical image gradient. To calculate the numerical image gradient for vertical direction ($\theta = 90$), the following equations can be applied

$$G_y^1(i,1) = I(i,2) - I(i,1) \tag{2.46}$$

$$G_y^N(i,N) = I(i,N) - I(i,N-1) \tag{2.47}$$

$$G_y^{INT}(i,j) = 0.5 \times I(i,j+1) - I(i,j-1), \;\; j \text{ starts at } 2 \tag{2.48}$$

where G_x^{INT} is the interior section of the numerical image gradient (blue in Figure 2.13) and G_x^1 and G_x^N are the first and last columns (yellow and red in Figure 2.13) of the image, respectively. A similar procedure can be applied to compute numerical image gradient for horizontal direction ($\theta = 0$). Therefore, the numerical image magnitude (N^{mag}) is computed using the following equation

$$N_{(i,j)}^{mag} = \sqrt{(I_{(i,j)})^2 + (I_{(i,j)})^2} \tag{2.49}$$

Other than that, a gradient feature for each of the pixel can also be computed using operators (e.g. Sobel, Kirsch, Roberts Cross, etc.) by applying a 3×3 convolution kernels over the entire image. Figure 2.14 shows an overview of feature extraction using 3×3 convolution kernels (horizontal, vertical and diagonal directions) of Sobel operators. To extract gradient at different orientations, each input image is convolved

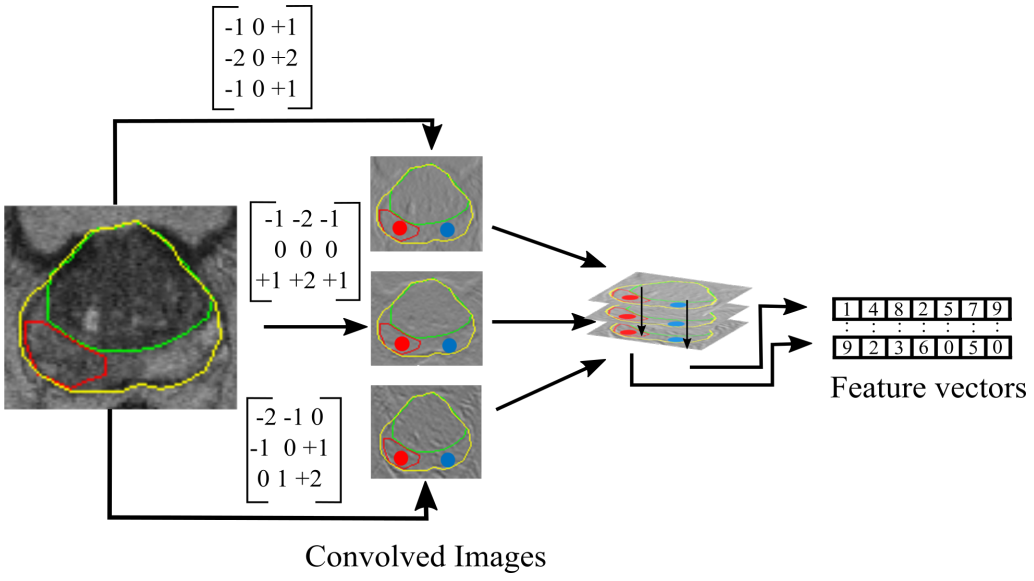

FIGURE 2.14: An overview of feature extraction using a 3×3 convolution kernel of Sobel operators.

separately with different kernels. For example horizontal kernel to detect gradient at $\theta = 0°$, vertical kernel to detect gradient at $\theta = 90°$ and diagonal kernel to detect gradient at $\theta = 45°$. Therefore, the magnitude in each pixel can be computed using the following equation

$$I^{mag}_{(i,j)} = \sqrt{(I^{\theta=0°}_{(i,j)})^2 + (I^{\theta=90°}_{(i,j)})^2} \tag{2.50}$$

where $I^{\theta=0°}_{(i,j)}$ and $I^{\theta=90°}_{(i,j)}$ are the pixel values at position (i, j) in image magnitude $\theta = 0°$ and $\theta = 90°$, respectively. The resulting image will be the image features.

2.4.7 Tamura's Features

Tamura *et al.* [174] developed a set of texture features which correspond to human visual perception. They defined six textural features namely contrast, directionality, coarseness, line-likeness, regularity and roughness. Their experimental results suggested that only the first three have significant effects in psychological measurements for human subjects (contrast, directionality and coarseness). Therefore, many studies have used only the first three features. Tamura contrast measures the variation of grey levels within an image (or a region) and to what extent their distribution is biased to black or white [175]. Tamura contrast can be computed using the following equation.

$$T^{cont} = \frac{\sigma}{(kur)^n} \tag{2.51}$$

where σ^2 is the variance and Tamura suggested $n = 0.25$. On the other hand, Tamura directionality (T^{dir}) is computed using the following steps [176].

- Step 1: Convolve (using 3×3) an input image (I) with horizontal and vertical Sobel edge operator. Each pixel has outputs $\bigtriangledown H_{(i,j)}$ and $\bigtriangledown V_{(i,j)}$.
- Step 2: Calculate the gradient vector for each pixel using the following equation

$$\bigtriangledown G_{(i,j)} = \frac{\bigtriangledown H_{(i,j)} + \bigtriangledown V_{(i,j)}}{2} \tag{2.52}$$

- Step 3: Calculate the orientation vector for each pixel using the following equation

$$\theta = tan^{-1} \frac{\bigtriangledown V_{(i,j)}}{\bigtriangledown H_{(i,j)}} \tag{2.53}$$

- Step 4: Construct a histogram (H_D) frequency versus θ, where $freq$ is the number of pixels with θ orientation (e.g. 5 pixels with $\theta = 45°$).

- Step 6: Identify peaks and valleys from H_D. Let N_p be the number of peaks identified (hence p is a peak). For each p, let w_p be a set of bins from its previous valley to its next valley, and ϕ_p is the angular position of the peak. In addition, h_ϕ is the height of the bin at angular position ϕ.

- Step 7: The directionality of an image is calculated as follow

$$\sum_{p}^{N_p} \sum_{\phi \in w_p} (\phi - \phi_p)^2 H_D(\phi) \tag{2.54}$$

Tamura coarseness (T^{crs}) is computed by applying the following steps [176].

- Step 1: For each pixel $I_{(i,j)}$ compute six averages of the neighbouring values at six different window sizes.

- Step 2: For each window (W) calculate the absolute difference (AD) between the pixels at the corners (diagonal directions). It can be done by subtracting pixels on the opposite side. Each scale has two AD values.

- Step 3: Select the W which has the largest AD.

- Step 4: The T^{crs} for each pixel in an image is the average of the neighbouring values which has the maximum AD. It can be computed using the following equation

$$\frac{1}{m \times n} \sum_{i}^{n} \sum_{j}^{m} W_{max}(i,j) \tag{2.55}$$

2.4.8 Fractal Features

Many textures have a statistical quality of roughness and self-similarity at different scales. Fractal based features show strong correlation with human visual perception at different resolutions, hence are suitable to characterise textures which appear at different scales [126, 141]. However fractal features are unable to differentiate similar textures at different orientations making them less suitable for describing textures with different orientations [152]. Although there are many ways to calculate fractal dimension (FD) features [126, 173, 177], all of those methods share a fundamental

principal to measure the roughness of the texture. The FD of an image can be measured by calculating the relationship between the different squared sizes and the overall area measured (the number of pixels). To further clarify, the FD of an image can be calculated using the following steps:

- Step 1: The input image is converted to a binary image.
- Step 2: The binary image is divided into a grid of squares (boxes) $k \times k$.
- Step 3: Count the number of boxes (nb) containing at least one pixel which is the '1' since it is a binary image.
- Step 4: Vary the size of the box to get different nb.
- Step 5: Plot a graph of $log(nb)$ versus $log(k)^{-1}$.
- Step 6: The curve is then approximated by a straight line using a line fitting method.
- Step 7: The FD of an image can be calculated by computing the slope of the line.

Alternatively, studies in [165, 166] extracted more FD features from an image by applying several different threshold values on the binary image. Subsequently, each thresholded binary image is processed by applying step 2 until step 7.

2.5 Summary

In summary, we have briefly discussed some of the fundamental concepts in computer vision, particularly image analysis, which will be the fundamental components in this thesis. From this chapter we are able to conclude that:

- Texture analysis is a powerful approach in pattern recognition which is consistently used in biomedical image analysis.
- Texture for the inhomogeneous category (texture with a random pattern) is among the most challenging tasks in texture analysis.
- Texture contains properties which can be computed statistically.

- Texture can be represented in many different ways such as histograms, shapes, vectors, curve lines, etc.

Chapter 3

Computer Aided Detection Systems

In prostate cancer imaging, CAD systems can help radiologists to identify abnormalities, by delineating those regions with the highest probability of being malignant. The MRI technologies have been integrated into the clinical practices and became part of the standard procedures in most hospitals around the world. This chapter aims to provide a brief review of the state-of-the-art of CAD systems for prostate cancer detection over the recent decades, focusing on the different stages within the CAD framework. These stages include all the vital steps such as pre-processing, multimodality image registration, prostate segmentation, feature extraction and selection and classification. The limitations of the current CAD systems in prostate cancer are then discussed as well as challenges.

3.1 CAD Status

MRI technologies have become crucial in the investigation of the human body anatomy. In the last 30 years, physicists have been focusing on improving imaging techniques assisting radiologists to improve cancer detection and diagnosis. However, a large number of MRI data need to be analysed affects the radiologist's performance and the ability of human visual diagnosis suffers variability causing inconsistent results. Another limitation is the lack of visibility due to the complexity of the clinical data such as overlapping structures in benign and malignant tissues making it impossible for human to analyse the image [178, 179]. Furthermore, unbalanced number

of cases between benign and malignant cases make human perception tends to follow the dominant structures resulting false positives and negatives [62, 178, 179]. In fact, substantial expertise from the radiologist is required to interpret multiparametric MRIs and such expertise is not widely available [46]. CAD systems can provide assistance with these problems either as a 'second opinion' to assist radiologists in making diagnosis decisions or as a 'first opinion' to automatically detect abnormal regions with minimal radiologists intervention. Computer algorithms allow the combination of large amounts of MRI data into fewer images with a compact information content, which can be manipulated to provide morphological and functional information about the organ [180].

To date, MRI prostate imaging is used to non-invasively diagnose tumor staging and improve treatment planning [181] rather than being used as a primary screening procedure to diagnose or detect prostate cancer. CAD systems for prostate cancer detection (CAD-PC) are relatively new in comparison to the other fields such as mammography [182] and CT colonography [183] due to the only recent development of endorectal coils which are improving image resolution. The use of different modalities in MRI such as T2-W, DCE, MRS and DWI were only developed after the mid 90's which makes these technologies relatively new. Interpreting prostate MRI requires a substantial amount of experience, is time consuming and often suffers from interobserver variation [184–186]. CAD-PC can speed up the reading time, minimise radiologist intervention and offer more consistent diagnosis results. In fact, the use of CAD-PC can enhance the diagnosis performance of radiologist significantly [187, 188]. For example, less experienced radiologists can achieve similar diagnosis performances to more experienced radiologists when assisted with CAD. Therefore, the development of CAD-PC is an active research topic taking MRI to the forefront of prostate cancer screening used in conjunction with the established clinical methods.

3.2 CAD-PC Framework

According to the most recent CAD-PC review conducted by Lemaître *et al.* [62] in 2015, 42 CAD-PC methods have been developed between 2007 and 2015. Although these methods used different approaches, there are some common components in a typical CAD-PC such as noise removal, feature extraction and classification (training and testing).

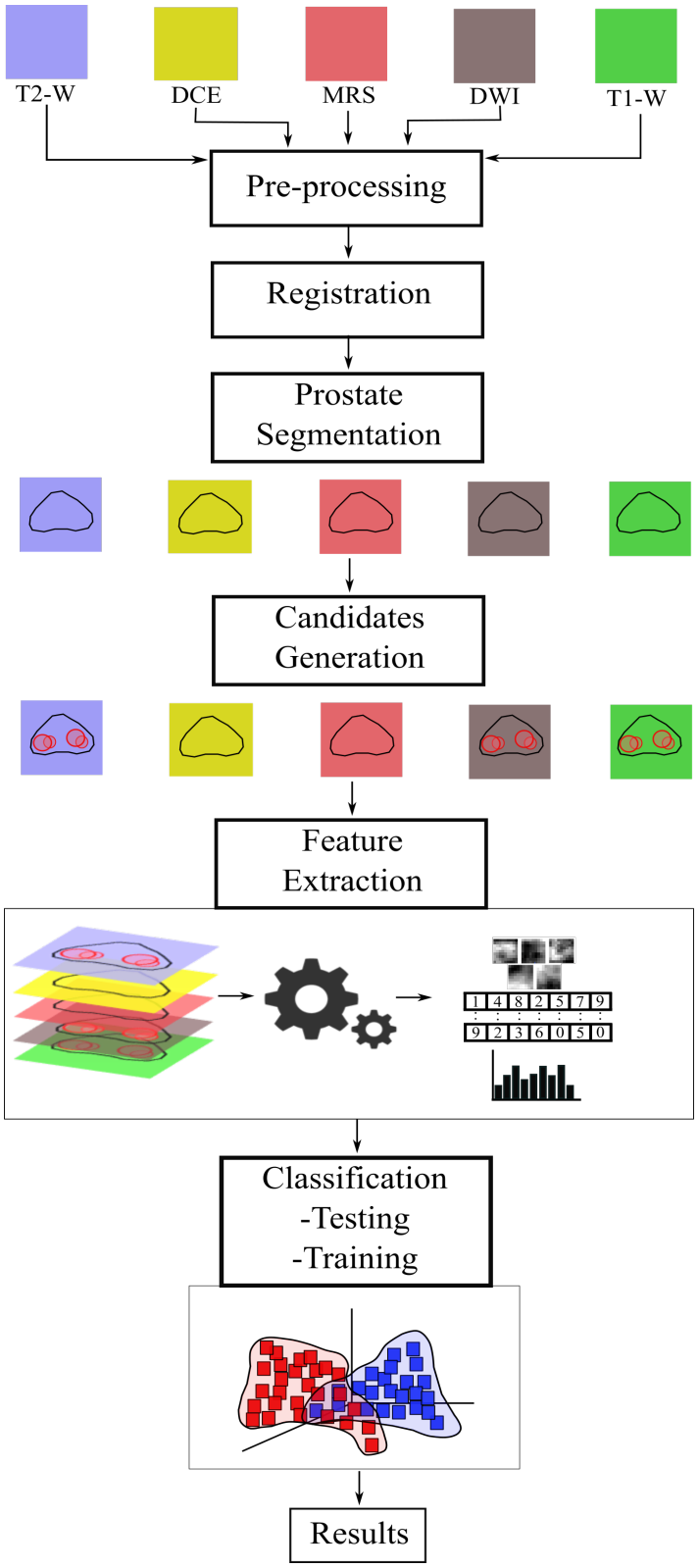

FIGURE 3.1: A general workflow of a typical CAD-PC.

Figure 3.1 shows a general workflow of a typical CAD-PC. For CAD-PC based on single modality consists of a pre-processing step, prostate and zonal segmentation (e.g. PZ or CZ), feature extraction and classification. Some CAD-PCs involve a 'candidates generation' step which means only certain regions of interest (ROI) will be further analysed in the next phase. On the other hand, CAD-PC based on multiparametric MRI generally contains an image fusion step called 'image registration'. In this step, MRI images from different modalities are combined into one single image to give a better picture of the anatomical structures of the prostate. However, some multiparametric CAD-PCs did not perform image registration but extract features from each MRI modality separately and combine all features into a single feature vector. Finally, different validation models were used in CAD-PC studies which will be discussed later.

3.2.1 Pre-processing

The major problem in MR images is that intensities do not have a fixed tissue specific numeric meaning even within the same MRI protocol, the same body region, and the same scanner [189–191]. These problems are mainly caused by [189–192]: a) corruption by thermal noise due to receiver coils, b) intensity variations due to different scanning protocols and c) poor radio frequency coil uniformity. Intensity variations in MR data can significantly affect performances of many image processing techniques, hence, they need to be corrected [191]. In the pre-processing step, each MRI image is either normalised, noise is removed or both.

In image normalisation the image is transformed to an image where all the pixel values are standardised across acquisitions. The main reasons for this are to make sure similar tissues have a common interpretation across locations and preserve the rank of intensities, hence the representation of the structure anatomy is less influenced by noise [193]. Indeed, one of the main challenges is the intensity variation between patients due to different acquisition settings. In one of the most popular normalisation techniques, image is normalised to zero mean unit variance. A study in [194] divides the original intensity by median + 2 × (inter-quartile range). Similarly, normalisation can be done by dividing each pixel intensity with 'standard z-score' of the PZ [195–198]. Shah *et al.* [199], performed normalisation by using the average fat signal near the prostate. Lv *et al.* [200] employed a method proposed in [201] which is based on

the assumption that MRI images from the same sequence share the same probability density function (PDF).

Another pre-processing step in CAD-PC is noise removal (or noise reduction). This step is crucial as signal intensity can be corrupted with noise or artifacts during the acquisition process. Many denoising methods have been developed in biomedical image analysis which can be found in [202, 203]. The simplest noise filtering method is median filter, but it does not work well on its own [62]. Studies in [192, 204, 205], suggested that performing anisotropic diffusion [205] on a median filtered image showed a significant improvement of the signal-to-noise ratio (SNR). Another denoising method is based on wavelet-based filtering using thresholding techniques as employed in [206, 207]. Lopes *et al.* [208] employed the filtering technique proposed in [209] which is based on joint detection and estimation theory.

Finally, the pre-processing step also involves a bias correction step known as bias field (intensity inhomogeneity). A bias field is where an artifact resulting in a very high intensity signal causes a layer of smooth intensity variation on the image. It often happened in brain MRI, but rarely in prostate imaging. Nevertheless, a couple of studies have performed bias correction: (a) Viswanath *et al.* [200] employed an inbuilt function in the Insight Segmentation and Registration Toolkit (ITK) library based on parametric Legendre polynomial model [210] and (b) Lv *et al.* [211] corrected the intensity inhomogeneity by iteratively detecting the smooth intensity layer (or foreground) using generalised scale. Subsequently, the bias field function is estimated based on the second-order polynomial function [62].

3.2.2 Image Registration

Image registration is popular in CAD-PC based on multiparametric MRI. Image registration is a process of overlaying two or more images of the same scene taken at different times, from different viewpoints, and/or by different sensors via a geometric transformation in an iterative procedure approach [62, 212]. In CAD-PC, image registration can be seen as a procedure of aligning different MRI image modalities into a template image [62]. Details of image registration methods can be referred in a survey reviewed by Zitová and Flusser [212]. Majority of the image registration methods developed in the literature consist the following steps: feature detection, feature matching, transform model estimation and image resampling and transformation.

For all the CAD-PCs reviewed by Lemaître *et al.* [62], the majority of the systems based on multiparametric MRI used parametric transformation models based on affine and elastic transformations. The affine transformation procedures offer the flexibility managing rotations and translation whereas elastic transformation is robust in handling local distortions [62]. In image registration, similarity measure is vital to make sure the detected features in reference and sensed images can be matched [212]. In this application, the most popular similarity measure method is estimating the Mutual Information (MI) [213] between detected features. MI assumes homogenous region in the first modality should appear as homogenous although they have different intensity values [62]. Another important step in multimodality image registration is optimising the geometric transformation model. Iterative estimation methods such as L-BFGS-B quasi-Newton method [214] and gradient descent [215] are among the common ones. Interpolation methods are used to approximate the pixel value in every single point in the transformed image.

Several CAD-PC have performed image registration such as in [200, 216–218]. Giannini *et al.* [217] and Vos *et al.* [21] used similar registration frameworks, by first finding an affine transformation using MI and then employed an elastic registration to build the geometric transformation model using the same similarity measure. However, Giannini *et al.* [217] used gradient descent approach [215] to optimise the transformation model whereas Vos *et al.* [21] employed L-BFGS-B quasi-Newton method [214] for the same purpose. Similar studies by Viswanath *et al.* [200, 219] to correct the misalignment between T2-W and DCE images used an affine registration together with MI similarity measure.

3.2.3 Image Segmentation

Image segmentation is usually performed after image registration. In studies of CAD-PC, segmentation is used to delineate the prostate gland where studies within the PZ aimed to separate the CZ and PZ regions. Prostate segmentation is an important task from both engineering and clinical perspectives. From an engineering point of view, an accurate prostate segmentation ensures all prostate tissues are covered in the later stage of CAD-PC including an accurate prostate volume (PV) estimation. Recently, PSA levels have been modified to derive the PSA density by incorporating PV calculations to help guide clinical decisions [220]. However, the accuracy of this method relies on the accuracy of the PV estimate. Therefore, from a clinical point

of view, an accurate prostate segmentation helps the accuracy of the PSA density and clinical diagnosis decisions. Prostate segmentation is a challenging task as the prostate gland boundary with its surrounding tissues can be ambiguous due to noise, artifacts and intensity inhomogeneity, hence it can be difficult to locate. In fact, different protocols of MRI scans employed at different medical centers or institutes make different tissue appearances on MR images (e.g. some scans use endorectal coils and some do not use coil) [221]. Even among experienced radiologists, the variability in accuracy for manual segmentation can be quite large.

Despite being a difficult task, there are many prostate segmentation methods have been developed in the last decade. Nevertheless, due to its complexity only a few methods have achieved good results. In 2012, the Medical Image Computing and Computer Assisted Intervention (MICCAI) society organised a Prostate MR Image Segmentation (PROMISE12) grand challenge [220]. The results from 2012 until 2015 revealed 17 teams have taken up the challenge and only five teams achieved a score of more than 80 with the highest score 84.36 achieved by a method proposed by Vincent *et al.* [222]. Firstly, the method built a prostate model using a Minimum Description Length (MDL) Groupwise Image Registration method and then employed a multi start optimisation scheme to match the model to test images. The Emory team (paper is under review by the PROMISE12 committees) was placed second achieved 83.66 where the method in [223] achieved 83.49. Birkbeck *et al.* [223] developed a method using Marginal Space Learning (MSL) to align a statistical mesh model on the image. Subsequently, using spatially varying surface classifiers the mesh is then hierarchically refined to the image boundary. A semi-automatic method of ETHZ group (paper is under review by the PROMISE12 committees) scored 81.24. A method of Malmberg *et al.* [224] scored 80.66 placing fifth in the ranking using '*Smart Paint*' software which gives options to users to fully interact with the 3D segmentation. The 2D segmentation result is shown in 2D slices through the object.

The other segmentation methods can be referred in [225]. Although many teams have taken up the challenge, there is still space of improvement. The average score is 76.84 $\pm$ 4.95 and the majority of the methods have the score between 70 to 80. Atlas-based segmentation, statistical modeling and image registration are among the most frequent techniques employed.

3.2.4 Candidate Generation

In CAD-PC, candidate generation is a process to segment possible malignant regions, so they can be further analysed in the next stage of the methodology which means the classification results are based on the number of regions (e.g. a region is classified as being malignant or benign, such as studies in [226–228]) or based on pixels but only within the segmented candidates. However, in some studies [200, 206, 207, 217, 229], very pixel is considered as a possible lesion hence the output defines pixels as either malignant or benign. Therefore, the classification results are based on larger number of instances.

Vos *et al.* [226] captured blobs on the ADC map as possible malignant regions based on the signal intensity and diameter. Subsequently the blobs detected are then filtered based on their appearances and only those which met the criteria are considered as final candidates. A similar approach used by Litjens *et al.* [227] to detect initial candidates using multi-resolution Hessian blob detector on the ADC map, T2-W and pharmacokinetic parameters maps. Subsequently, feature extraction was performed on each detected blob before the SVM classifier was employed to generate final candidates. Later, Litjens *et al.* [228] modified the method by using local maxima detection to capture initial candidates followed by a post-processing step by employing region-growing and morphological operations. A set of features were then extracted from each of the candidates before k-NN classifier was employed to select the final candidates.

3.2.5 Feature Extraction

Feature extraction is an important task to capture information representing all textures in an image. This process can be done using several techniques such as DWT [230], Gabor Filters [162, 163], GLCM [106, 108], LBP [74, 75] and some other methods which have been discussed in Chapter 2. In this section, a number of features used in CAD-PC studies will be discussed. In CAD-PC, features are computed either based on the final candidates (a feature value represents each candidate) or pixel based where each feature value represents each pixel.

Since different modalities provide different features, the set of features in every method in the literature vary. For example texture and intensity based features are very popular in T1-W and T2-W MRI whereas parameter based features can be extracted from MRS or DCE modalities. Nevertheless there are features consistently used by most

methods such as intensity-based, edge-based, statistical-based and texture-based. According to the review conducted by Lemaître *et al.* [62], the most popular features are first- and second-order statistical features. In T1-W and T2-W MR images, the majority of the methods [206, 207, 217, 231–234] in the literature used Haralick features (GLCM), intensity-based and edge-based features (e.g. Sobel and Kirsch operators). Statistical-based features such as percentiles and Statistical-moments are also among the most popular methods whereas histogram-based features such as LBP, histogram orientation gradient (HOG), Wavelet-based, Gaussian filter bank and PDF are among the least popular features. Other features used are fractal, position and anatomical-based features.

In contrast, DWI, DCE and MRS modalities provide different sets of features. In DWI, the ADC map is usually used as a source for blob detection. Studies in [226, 228] used the ADC map to capture possible malignant regions before they were further analysed. On the other hand, in DCE modality, a set of features represent levels of different metabolites in prostate tissues such as wash-in and wash-out rates, absolute maximum enhancement, 95% of maximum enhancement, onset time, time to peak, time to max (T_{max}) and the area under the gadolinium curve (also known as semi-quantitative features) are among the most popular features used. Quantitative features were also extracted such as forward volume transfer constant (K_{trans}), the fractional volume of extracellular space per unit volume of tissue (ve) and the reverse reflux rate constant between extracellular space and plasma (kep). In MRS modality biological markers such as choline, creatine and citrate metabolites are mainly used in the literature. All these features are either extracted on the basis of pixel-based or region-based approaches (depending on the method). Subsequently, all features are incorporated into the feature vector for classification.

3.2.6 Feature Selection

Feature selection is a process to select only relevant information, hence eliminating unnecessary data and reducing the data dimensionality. In other words, only features which have significant correlation with classes (e.g. benign and malignant) will be considered in building the predictive model. Feature selection is necessary to a) avoid over-fitting when building a classifier model as less data means less chance of making decisions based on noise, b) possibly improve accuracy because only the most relevant attributes are selected to build the predictive model (a study conducted by Niaf *et al.*

[231] demonstrated that feature selection significantly improves the discrimination performance between malignant and benign regions) and c) reduce training time because fewer features are used in making decisions. The most popular feature selection methods employed in the study of CAD-PC are based on t-test, mutual information and minimum-redundancy–maximum-relevancy criteria. In most cases, before feature selection is performed, each feature is usually normalised to avoid that absolute values playing a role [235].

3.2.7 Classification

The final step of a CAD-PC is classification which involves training and testing with features extracted from images. In the training phase, a predictive model is built by the selected classifier usually by learning the labelled data. Subsequently, in the testing phase the built model is applied to the test data without the knowledge of ground truth labels.

In the past two decades, studies have used many different classifiers and results vary even tested on the same dataset and methodology. Regardless of the choice of kernel (e.g. RBF, polynomial and Gaussian), the SVM classifier is the most popular supervised learning method used in the study of CAD-PC. However, one of the disadvantages of SVM is the performance is extremely dependent on the parameter settings. To tackle this issue, a genetic algorithm was employed in which each SVM hyperparameter was treated as a "gene" in a "chromosome" encoding and the F-measure was used as the metric for the evaluation of individual fitness [221]. For linear model classifiers, the linear discriminant analysis (LDA) is the second choice [46, 226, 233, 236] and the logistic regression is only used by the studies in [237, 238].

In recent years, the random forests classifier was introduced to the medical image analysis community and achieved very promising results in some medical applications [239]. One of the advantages is the fact that it can handle high dimensional data as well as large number of training samples [240]. In addition, its ability to estimate what features are important in the classification is another attraction of random forests. On the other hand, clustering methods such as k-means clustering and k-NN are also quite popular. These methods are very easy but can work quite well which makes it very attractive. Adaboost and multilayer perceptron are among the learning methods which are recently used in CAD-PC.

3.2.8 Model Validation

In any applications of CAD, the validation model for accessing the performance of the classifier model is crucial. In CAD-PC, two techniques are broadly used: (a) Leave-one-out Cross-validation (LOOCV) and (b) n-fold Cross-Validation (n-CV). In LOOCV, data from one patient are kept for validation and the other data from the other patients (all patients except the one kept for validation) are used for training. This validation technique is popular when dealing with small number of patients, hence considered unreliable [241]. On the other hand, n-CV was used by most studies with a large number of patients. The n-CV technique splits dataset into n subsets where the samples are randomly selected. For example, if the dataset contains 45 patients, and $n = 9$, then each subset has 5 patients. This means, all data from the other 40 patients will be used to build the classifier model and data from the five patients will be used for testing. This technique is regarded as a more reliable model validation [241].

3.3 Open Challenges in CAD-PC

Developing a robust CAD-PC remains a challenging task due to the following problems:

- The state-of-the-art of prostate gland segmentation accuracy is still quite low with an average score 76.84 ± 4.95. The best method in the literature achieving below 85.
- The MRI images qualities can be vary depending on the protocols used by different scanners. Hence, more accurate and standardised scanning procedures are needed.
- It is impossible to make quantitative comparisons among the CAD-PC due to the absence of public data. Therefore, it is difficult to determine whether one method works well on a different dataset or not.
- Most of the proposed methods in the literature were tested on a small dataset leading to uncertain results when tested on a larger dataset.

- The appearance of malignant regions are still uncertain. In T2-W MRI, a region dominated by low intensities has a higher chance of being malignant. However, there are other factors which can cause this appearance and lead to false positives.

It is therefore expected that an improvement in one or more of these problems will lead to a better diagnostic system with higher sensitivity and lower false positive rate, as has been observed from previous research efforts. It is likely that more improvements will be made in the next decade and wide deployment of prostate CAD systems in the clinical environment will eventually occur.

3.4 Summary

This chapter has summarised the workflow overview in CAD-PC based on the existing methods in the literature which covered all of the methods reviewed by Lemaître *et al.* [62]. From this chapter, we can see that there a typical CAD-PC has few similarities:

- Normalisation is seen to be a crucial step to standardise the representation of tissues intensities across patients. This is similar to a noise removal step to ensure data is less effected by noise or artifacts.
- An accurate prostate gland segmentation is crucial for studies involved in prostate volume estimation. Nevertheless, studies in prostate cancer detection focus more on the classification results rather than the accuracy of the segmentation of the prostate boundary.
- CAD-PC studies based on multiparametric MRI performed image registration for better anatomical and feature representation.
- Using a machine learning technique to build a predictive model for classification is the main trend in CAD-PC.
- There are many texture descriptors have been used in CAD-PC studies and first- and second-order statistical features are the most popular.

- Studies of CAD-PC based on multiparametric MRI are in favor among researchers because it can provide morphological and functional information about the prostate. Nevertheless, CAD-PC studies based on single modality are still relevant.

Chapter 4

Detection and Localisation of Prostate Cancer using Intersection of Binary Segmentations

This chapter proposes a CAD-PC methodology based on combining multiple segmentation techniques. We extract four image features using Gaussian and median filters. Subsequently, we use each image feature separately to generate binary segmentations. Finally, we take the intersection of all four binary segmentations, incorporating a model of the peripheral zone, and perform erosion to remove small false positive regions. The evaluation of this method is based on 275 T2-W MRI images from 37 patients and 86% of the slices were classified correctly with 87% and 86% sensitivity and specificity achieved, respectively. This chapter makes two contributions: firstly, a novel Computer Aided Diagnosis approach which is based on combining multiple segmentation techniques using only a small number of simple image features. Secondly, the development of the proposed method and its application in prostate cancer detection and localisation using a single T2-W MRI modality with the results comparable to the state-of-the-art multimodality and advanced computer vision methods in the literature.

4.1 Introduction

The ultimate goal of this chapter (and thesis) is to develop a CAD tool for prostate cancer detection and localisation within the PZ mainly because a) about 80% of the

prostate cancers appear in the PZ [12, 192, 242, 243] and b) in general, prostate cancer that arises in the peripheral zone is more aggressive than that which arises in the transitional or central zones [244] (as discussed in Chapter 1). Therefore, in this paper we propose a new method for detecting prostate abnormality within the PZ using four different image features (see Figure 4.5) extracted using Gaussian and median filters. The main objective of this method is to delineate malignant regions (and hence localise them) within the PZ by taking the overlapping binary segmentation from each image feature. This means pixels (or tissues) which are classified in the same malignant cluster in all image features are considered to have the highest probability of being malignant. However, if a pixel is classified as belonging to a benign or normal tissue cluster in one of the image features we considered it to be a benign or normal tissue (this will be explained in more detail in section 6.3). Figure 4.1 illustrates the basic idea behind the proposed method in this chapter. The motivation behind this idea is that each texture descriptor represents different information about the texture content. Therefore, segmenting the image feature individually resulting to different binary segmentations. By finding the overlapping binary segmentations we are able to find the region with the highest probability of being malignant. In other words, the pixels within this region are clustered as malignant in feature representation.

The novelty of this method resides in an approach which combines simple features (this is similar to a forest of weak classifiers which together provide strong results), for the first time and applied to prostate T2-W using one modality. To our knowledge, no existing methods in the literature have used the technique of finding cancer regions by

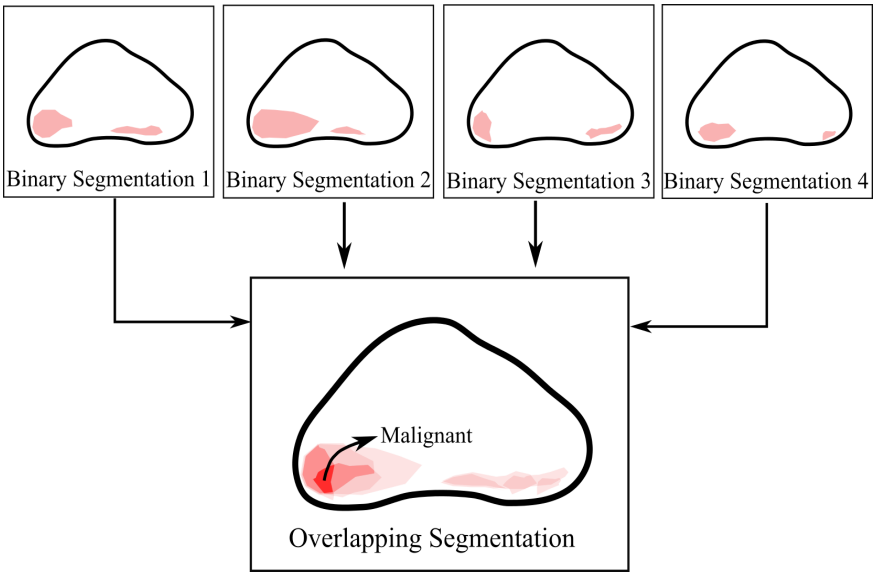

FIGURE 4.1: An illustration of the basic idea behind the proposed method.

taking the overlapping binary segmentation extracted from a small number of image features.

4.2 A Brief Review of Segmentation Techniques

CAD can be applied in many different medical imaging applications (such as brain, breast, chest and prostate) using different segmentation techniques. As previously mentioned in Chapter 2, there are six different categories of segmentation technique in CAD systems namely contour and shape based, machine learning based, region based, statistical based, multiresolution based analysis, hybrid and soft computing methods. The contour and shape based methods are the most popular techniques employed in the literature. The other techniques and their applications in medical imaging can be found in [236, 245–249].

The level set technique has been applied to several human organs (e.g. brain, cardiac, prostate, breast, etc). Tsai *et al.* [250] developed a shape-based approach using level sets and demonstrated their method by applying it to the segmentation of cardiac and prostate MRI. The proposed method derived a parametric model for an implicit representation of the segmenting curve by applying principal component analysis to a collection of signed distance representations of the training data. The parameters of this representation were then manipulated to minimize an objective function for segmentation. Liu *et al.* [251] proposed a method for mass segmentation in mammograms using a level set to improve the initial segmentation performed using a watershed algorithm. On the other hand, Shi *et al.* [252] used k-means clustering followed by a morphological opening operation for initial mass segmentation. For the level set segmentation a linear discriminant analysis (LDA) classifier with stepwise feature selection was used to merge the extracted features into a classification score.

In [253], a novel automatic approach to identify brain structures in magnetic resonance imaging (MRI) is presented for volumetric measurements. This approach combines the active contour model with a support vector machine (SVM) classifier. The SVM features are selected according to the structure of brain tissues: gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF). Jones *et al.* [249] developed an interactive segmentation method by combining both region selection and user point selection. Evaluation results showed an accuracy on average of more than 98% based on 248 intravascular ultrasound (IVUS) images.

Graph searching techniques have been studied for the segmentation of biomedical images such as brain, knee and glottis [87, 88, 254]. Pedoia and Binaghi [87] proposed a fully automatic 2D brain segmentation using graph searching technique which consisted of border detection based on two-dimensional graph searching principles and radial contour detection. In the border detection phase, polar conversion is performed first, followed by skull and brain boundaries detection. Li *et al.* [88] extended the optimal graph-searching techniques to 3D and higher dimensions using a polynomial-time algorithm for surface segmentation in volumetric images. The method is efficient and robustly tested on MR arterial walls and IVUS image. Finally, a study in [254] uses the output of SVM to drive a graph-cuts segmentation, which was initially trained as a local Golgi detector based on rotationally invariant features. It should be noted that in some of the described segmentation approaches difference images have been used but in general these rely on multimodality data or on some temporal sequences, neither of which have been used in the proposed approach. In the developed approach we rely on a single modality and the variation in appearance between normal and abnormal tissues within the peripheral zone.

Despite the relevant achievements obtained, the main limitations of contour and shape based methods for our research are

1. Most of them require a significant amount of user interaction for initial region selection. In our case, we want to eliminate (or minimise) user interaction in finding cancerous regions.

2. They work well only if the boundary of the object is well defined within the image: such methods work well for prostate gland detection [255]. In many prostate T2-W MRI, cancerous regions are vague both in terms of appearance and shapes, and as such many training samples would be needed.

3. The results of many of these methods are highly dependent on the initialisation.

Therefore, we applied an unsupervised machine learning based method which is Fuzzy C-Means (FCM) clustering as proposed by Chen and Zwiggelaar [256]. Motivated by the advantages and disadvantages of the range of FCM based algorithms the authors proposed a modified FCM (mFCM) algorithm that incorporates local spatial and intensity information based on an adaptive local window filter whose weighting coefficients differentiate the neighbouring pixels within the local window. The method

is less sensitive in dealing with different types of noise, intensity inhomogeneities, excludes outliers, and eliminates unnecessary blur. Moreover, the separation of filtering from clustering can drastically reduce the running time. The mFCM has a two-pass filtering process: firstly is to evaluate to distinguish between unreliable and reliable neighbouring pixels and secondly a new intensity value of the central pixel is computed based on the reliable neighbours values to generate the filtered image. Subsequently, a fast clustering is performed on the filtered image using intensity histogram. The mFCM algorithm will be described in details in section 1.4.4.

4.3 Modelling the Peripheral Zone

Figure 4.2 shows an example MRI image with the ground truth of the prostate gland, central zone and tumor represented in yellow, green and red, respectively (left image), while the right image shows a simpler schematic overview of the prostate derived from the prostate anatomy proposed in [257] with CZ, PZ, and tumor (T).

Pathologically, about 80% of prostate cancers arise in the PZ and the rest are within the CZ [12]. Since the percentage occurrence of cancers in the PZ is high and as these tend to be more aggressive, we aim to detect prostate abnormality within this region. We did not perform prostate segmentation because all prostates were already delineated by an expert. It should be noted that Zhu *et al.* [255] developed a method to detect the prostate capsule. Based on the schematic overview shown in the right image of Figure 4.2 (also proposed in [257]), we defined our 2D prostate model based on Figure 4.3.

The generic prostate's PZ model in this paper is mainly inspired from similar models proposed by Makni *et al.* [258] and Liu *et al.* [259] which used catenary and polynomial curves, respectively. From a radiological point of view, the prostate is

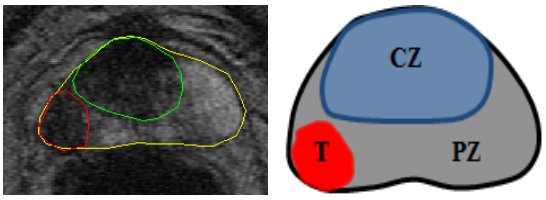

FIGURE 4.2: Prostate MRI image (left image) with its ground truth delineated by an expert radiologist and a schematic (right image) overview of the prostate containing a tumor.

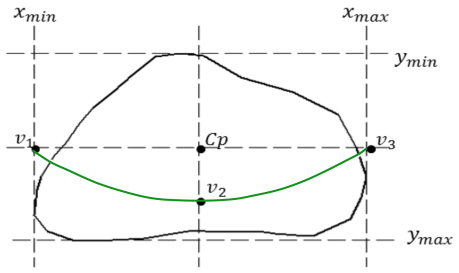

FIGURE 4.3: Prostate gland (black) and the defined PZ below $y = ax^2 + bx + c$ (green) which goes through v_1, v_2 and v_3.

mainly divided into two regions in MR (the PZ and CZ). Therefore, according to Makni *et al.* [258], when segmenting these regions, expert radiologists tend to follow the rule of 'imagining' outlines due to contrast or strong artifacts. The process of distinguishing these regions is heavily relying on a *priori* knowledge of their most likely locations [258]. Indeed, a more accurate way (probably more complex and time consuming) could be achieved by segmenting the PZ within the prostate gland. However, these approaches are complicated and require high accuracy in distinguishing tissues in the PZ and CZ. In cases where there is no clear boundary between the PZ and CZ, most segmentation based methods suffer from over-segmentation (hence, could lead to many false positives). In contrast, defining quadratic curves is simple and fast.

In our CAD system, we used the quadratic equation $y = ax^2 + bx + c$ based on three crucial coordinate points of the prostate which are v_1, v_2 and v_3, which are determined by the outmost x and y coordinates of the prostate boundary: $x_{min}, x_{max}, y_{min}, y_{max}$ (see Figure 4.3). For example, x_{min} and y_{max} can be determined by taking the minimum x and maximum y coordinates along the prostate boundary. Moreover, the x coordinates of v_1 and v_3 are captured from x_{min} and x_{max} and their y coordinate is determined by taking the y coordinate between y_{min} and y_{max}. On the other hand, the x coordinate of v_2 is taken from the x coordinate x_{min} and x_{max} and its y coordinate is determined by taking $\frac{3}{4}$ (0.75) of the distance from y_{min} to y_{max}. The coefficient ($\epsilon = 0.75$) is selected as it gives balanced results in terms of accuracy, sensitivity and specificity (see Figure 4.18). Mathematically, these can be represented in equations (4.1), (4.2), (4.3) and (4.4).

$$C_p = ((x_{min} + x_{max})/2, (y_{min} + y_{max})/2) \tag{4.1}$$

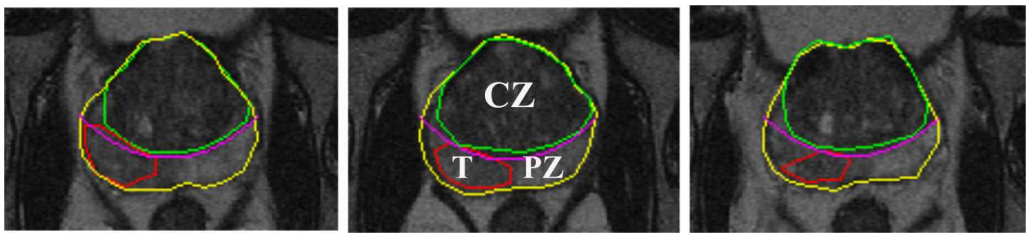

FIGURE 4.4: Three different prostate MRI images with ground truth delineated by an expert radiologist and the estimated PZ region indicated under the magenta line.

$$v_1 = (x_{min}, (y_{min} + y_{max})/2) \tag{4.2}$$

$$v_2 = ((x_{min} + x_{max})/2, y_{min} + ((y_{max} - y_{min}) \times \epsilon)) \tag{4.3}$$

$$v_3 = (x_{max}, (y_{min} + y_{max})/2) \tag{4.4}$$

Once the coordinates of v_1, v_2 and v_3 are defined, we can determine the values of a, b and c (therefore a final quadratic equation is defined). Finally, by taking every x coordinate from x_{min} to x_{max} into a quadratic equation we are able to determine the y coordinate which will define the PZ's boundary (the main goal is to analyse the region under the green line in Figure 4.3). The approximation model is able to capture most of the PZ area, easy to implement and computationally efficient. Figure 4.4 shows examples of the estimated PZ boundaries in three different MRI images produced by this method.

4.4 Methodology

Figure 4.5 shows the overview of the proposed methodology. First, we perform Gaussian and median filtering on the original image to obtain G_1 and M_1. We extract a probability image from G_1 and M_1 using greyscale frequency before we obtain the third feature (F_1) which is the magnitude of G_1 and M_1. This means, each element in F_1 is the sum vector of each component from G_1 and M_1. On the other hand, the fourth feature (F_2) is the vector magnitude of probability images from G_2 and M_2. Subsequently, we use each feature separately to generate binary segmentations taking the feature space and intensity values into account. We perform erosion on each of the segmentations to remove small false positive regions. Finally, we take the intersection of all four binary segmentations, taking a model of the peripheral zone (see Section 3) into account.

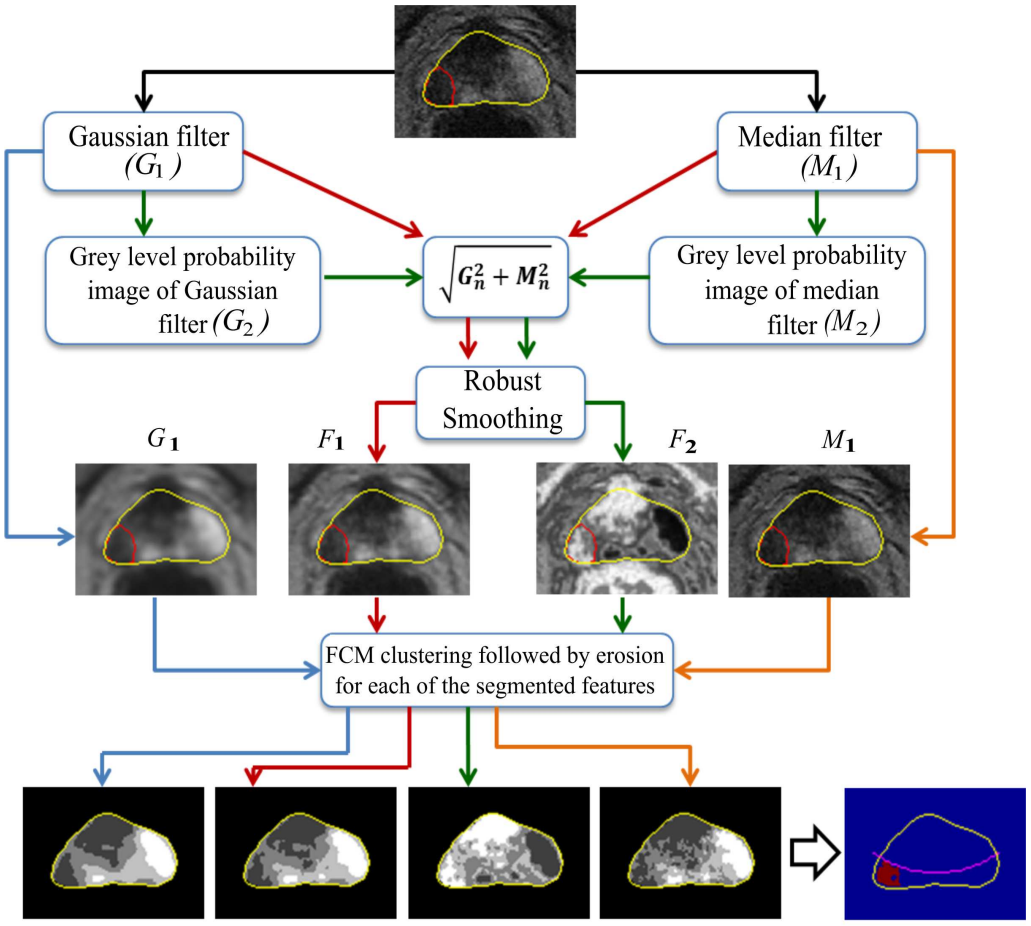

FIGURE 4.5: Overview of the proposed methodology. Intensity values are represented by greyscale colours, with the darkest representing the lowest intensity.

4.4.1 Pre-processing

Since MRI images often suffer from different types of noise, it is necessary to apply different types of denoising methods before doing any further processing. From a clinical point of view, this is an appropriate step to enhance characteristics of an important region of interest (such as textures and boundaries). Moreover, it does not deform the anatomical locations of tissue regions because this step deals only with noise without affecting the spatial information.

Hendrick [260] reported that one type of noise in MRI images is Gaussian noise. Therefore, Gaussian smoothing is selected, which is also effective in reducing noise that is problematic for image analysis algorithms. For example, image segmentation is often affected by the presence of too many local minima/maxima and inflection points in the data [261]. Studies performed by Barentsz *et al.* [262] and Viswanath *et al.*

[263] suggested that most cancers shows textural distortions in T2-W images. Litjens *et al.* [46] captured these characteristics using Gaussian filters. Many previous studies have applied Gaussian filters in denoising MRI images and mammograms [260, 264, 265]. We are aware that there are more sophisticated denoising methods such as those based on Fourier analysis [266] or anisotropic filtering [267] which could be explored in future work. The 2D Gaussian function is defined as

$$g(s,t) = \frac{1}{2\pi\sigma^2}.e^{-\frac{s^2+t^2}{2\sigma^2}} \tag{4.5}$$

where s is the distance from the origin in the horizontal direction, t is the distance from the origin in the vertical direction, and σ is the standard deviation of the Gaussian distribution. In the proposed method we used the following parameters: the kernel size (ks) is 15×15 and the standard deviation (σ) is 3.0. See subsection 4.4.3 for the selection of the Gaussian parameters (σ and ks) and their resulting variability can be seen in Figure 4.17.

On the other hand, we used median filtering to preserve the regional boundaries (e.g. tumor regions). It is claimed that using median filtering is much better at preserving sharp edges [268] and in our case we want to preserve the information-bearing structures such as tumor boundaries [192]. The median filter works by replacing the pixel value with the median value in the neighborhood of that pixel. We used a sliding window of 5×5 pixels. Other sizes are possible (such as 3×3, 7×7 and 9×9) give similar results.

4.4.2 Feature Extraction

In the previous step we have extracted two texture features using Gaussian (G_1) and median (M_1) filters. In this section, two more texture features will be extracted by manipulating G_1 and M_1. We calculated the probability images using equation (4.6) and calculate the vector sum using equation (4.7). Probability images are commonly used to model the expected appearance of an object (e.g. tumor region) in a given reference space. Many studies suggested that prostate cancer appears darker within the PZ and is similar to the tissues outside the prostate gland. By computing a probability image we are able to quantify the likelihood of every pixel/voxel belonging to specific tissues (e.g. tumor region). This also means each pixel/voxel can be represented by a value of the likelihood of being malignant. The conversion to probability

images acts as a normalisation across the feature images and as such these can all be treated in the same framework. We calculate the probability image for each of G_1 and M_1. This means, for an $f(i,j)$ image, the probability value for the k^{th} grey level is calculated using:

$$P(i,j) = \frac{\#(f(i,j)=k)}{M \times N} \tag{4.6}$$

where $\#(f(i,j)=k)$ is the number of pixels at the k^{th} intensity level in a $M \times N$ image, and as such each element in P is the probability value for a particular intensity level. Others have exploited probability images for segmentation [269, 270]. To calculate features F_1 and F_2, we find the sum vector for every corresponding element in G_1 and M_1 using equation (4.7). According to [271] by taking the square root of the sum of squares between two corresponding signals produces a good image with little noise and continuous edge marking while another study performed in [272] suggested that equation (4.7) can improve the signal to noise ratio (SNR). In this case, corresponding signals are the corresponding pixels in G_1 and M_1.

$$I_n(i,j) = \sqrt{G_n^2(i,j) + M_n^2(i,j)} \quad , n = 1 \; or \; 2 \tag{4.7}$$

In total, four features are extracted, namely the Gaussian feature (G_1), the median feature (M_1), the magnitude Gaussian and median features (F_1) and the magnitude of the probability images of Gaussian and median features (F_2). Before image segmentation is performed, we applied noise reduction to F_1 and F_2 to minimise the noise retained/created after being processed using equations (4.6) and (4.7). In the proposed method we applied a smoothing method developed by Garcia [273] which is robust in dealing with weighted, missing, and outlying values by using an iterative procedure (which is the case in G_2 and M_2). The method is a fast and robust version of discrete smoothing splines based on discrete cosine transform (DCT). The algorithm, uses penalised least squares method, allows fast smoothing of data in one and higher dimensions. Moreover, the method automatically determines the amount of smoothing carried out by minimising the generalised cross-validation score, hence avoiding over- or under-smoothing [273].

Figure 5.5 shows examples of all extracted features G_1, F_1, F_2 and M_1. We can see that the malignant region appears brighter in F_2 and darker in the other features.

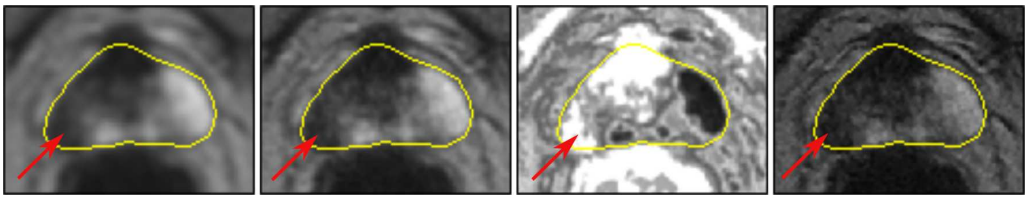

FIGURE 4.6: Example extracted features, from left to right: G_1, F_1, F_2 and M_1. The outline prostate is defined by the yellow line and the malignant region (in the PZ) is indicated by the red arrow.

4.4.3 Gaussian Parameters

The selection of the parameters for the Gaussian smoothing function is based on the studies in [274–278], which indicated that, the standard deviation (σ) and kernal size (ks), are linked. According to the experiments with different Gaussian convolution algorithms conducted in [274], the authors showed that the amount of error (the smaller the error the closer the denoised image is in comparison to the original image) did not change significantly after $\sigma \geq 2$. This means, for many Gaussian algorithms the error is much higher (less accurate) when $\sigma<2$. On the other hand, for the selection of kernel size (ks) several authors [275–278] suggested that in general, filter size should be $\lceil 3\sigma \rceil$ to $\lceil 5\sigma \rceil$ and odd [278, 279]. For instance, if $\sigma = 2.5$, the recommended minimum kernel size is 9 ($3 \times 2.5 = \lceil 7.5 \rceil = 8$, since it should be an odd number according to [278, 279], the nearest odd value is 9). Similarly, selecting $\sigma = 1$ would suggest the smallest kernel size of 3×3. Although there are no quantitative experimental results for optimal Gaussian parameters on medical images such as MRI or ultrasound, their results indicate a general guideline for selecting Gaussian parameters. In the proposed method we used several σ values together with several kernel sizes and chose the ones that give the highest accuracy, sensitivity and specificity (see Figure 4.17). The selection of parameters is not the major focus of this study but the development of a novel method of prostate cancer detection and localisation within the PZ is.

4.4.4 Clustering

In the proposed method, image segmentation is performed using a Fuzzy C-Means (FCM) algorithm as it has been widely applied in a variety of medical image segmentation applications [256, 280]. However, one common problem with FCM is its

ability in handling different types of noise and intensity inhomogeneities taking local spatial and intensity information into account. The conventional FCM algorithm [281] assumes that every pixel can belong to multiple classes with varying degrees of membership. The algorithm works by assigning membership to each data point corresponding to each cluster center on the basis of distance between the cluster and the data point. The closer the data point to the cluster center the higher its membership value for that cluster. Let $Y = (y_1, y_2, ..., y_R)$ denote an image with R pixels to be partitioned into d clusters. FCM iteratively minimises the objective function defined as

$$J_{fcm} = \sum_{p=1}^{d} \sum_{q=1}^{R} u_{pq}^{m} \|y_q - v_p\|^2 \tag{4.8}$$

with the following constraints: $\sum_{p=1}^{d} u_{pq} = 1$ for $\forall q$ and $0 < \sum_{q=1}^{R} u_{pq} < R$ for $\forall p$, where u_{pq} represents the membership of pixel y_q to the p^{th} cluster, y_q represents the feature data of the q^{th} pixel, and v_p is the prototype value of the p^{th} cluster centre. The parameter m (equal to 2 in this study as defined in [109]) is a weighting exponent on each fuzzy membership that controls the fuzziness of the resulting partition.

Motivated by the advantages and disadvantages of the range of FCM based algorithms, Chen and Zwiggelaar [256] presented a number of modifications to the Enhanced FCM (EnFCM) [282] and fast generalised fuzzy c-means (FGFCM) [283] algorithm, hence propose a mFCM algorithm which incorporates local spatial and greylevel information. An adaptive local window filter is introduced which differentiates the weight of the neighbouring pixels when computing the filter response value of the central pixel. Moreover, the mFCM algorithm automatically determines the weighting coefficients of the neighbouring pixels based on their spatial and greylevel associations with the central pixel. The mFCM algorithm works as follow [109]:

1. Define a local square window W centred on pixel $W_{(i,j)}^{c}$ (the surplus 'c' indicates central). The authors proposed a window size of 5×5, hence this size is used in this method (other shapes and sizes are possible). $W_{(i,j)}$ represents each of the pixels (except the central pixel) in W.

2. Find the median intensity (W_{med}) and compute the intensity deviation (W_{σ}) within the neighbouring pixel values in W.

3. Find the reliable and unreliable neighbouring pixels. Reliable pixels are those with difference value is larger than W_σ. The difference value for each neighbouring pixel can be computed by subtracting $W_{(i,j)}$ with W_{med}. Let denotes $W^r_{i,j}$ as reliable neighbouring pixels (the surplus 'r' indicates reliable).

4. Define a local window filter and compute the weighting coefficient (C_{rc}) between $W^r_{i,j}$ and $W^c_{(i,j)}$.

$$C_{rc} = \begin{cases} C_{\Delta\hat{D}}.C_{\Delta\hat{I}} & \text{if } W_{(i,j)} \in W^r_{(i,j)} \\ 0 & \text{otherwise} \end{cases} \tag{4.9}$$

where $C_{\Delta\hat{D}}$ and $C_{\Delta\hat{I}}$ are coefficients determined by local spatial distance and the local intensity difference between $W^r_{(i,j)}$ and $W^c_{(i,j)}$, respectively. Both can be computed using the following equations:

$$C_{\Delta\hat{D}} = \begin{cases} exp\left(\frac{(-\Delta\hat{D})}{\lambda_{\hat{D}}}\right) & \text{if } W_{(i,j)} \in W^r_{(i,j)} \\ 0 & \text{otherwise} \end{cases} \tag{4.10}$$

$$C_{\Delta\hat{I}} = \begin{cases} exp\left(\frac{-(W^c_{i,j}-W^r_{i,j})^2}{\lambda_{\hat{I}}.\sigma^2_{rc}}\right) & \text{if } W_{(i,j)} \in W^r_{(i,j)} \\ 0 & \text{otherwise} \end{cases} \tag{4.11}$$

where $\lambda_{\hat{D}}$ and $\lambda_{\hat{I}}$ are the scales which influence the $C_{\Delta\hat{D}}$ and $C_{\Delta\hat{I}}$, respectively. In this study, we used the original parameters suggested by the author, $\lambda_{\hat{D}} = 1$ and $\lambda_{\hat{I}=3}$. The Euclidean distance between pixels $W^c_{(i,j)}$ and $W^r_{(i,j)}$ formulated $\Delta\hat{D}$, where σ^2_{rc} is the intensity deviation from the central pixel $W^c_{i,j}$ with each of the reliable neighbouring pixels ($W^r_{i,j}$) which can be calculated as follow

$$\sigma_{rc} = \sqrt{\frac{\sum_{W_{(i,j)} \in W^r_{(i,j)}} (W^r_{(i,j)} - W^c_{(i,j)})^2}{N^r}} \tag{4.12}$$

where N_r is the number of reliable neighbouring pixels.

5. The image is filtered using the defined local window filter. The resulting intensity value ($R^c_{i,j}$) of the central pixel is computed by:

$$R^c_{i,j} = \frac{\sum_{W_{(i,j)} \in W^r_{(i,j)}} \left(C_{rc}.W^r_{(i,j)}\right)}{\sum_{W_{(i,j)} \in W^r_{(i,j)}} (C_{rc})} \tag{4.13}$$

6. From the intensity histogram of the filtered image count the number of greylevels and the number of pixels with the same greylevel.

7. Cluster the filtered image based on its intensity histogram using the EnFCM algorithm [282].

In the proposed method, we segment every image feature, separately into four different classes based on the number of tissue categories in the prostate: normal (non-neoplastic) prostatic tissue, benign prostatic hyperplasia, high-grade prostatic intraepithelial neoplasia, and prostatic adenocarcinoma [284]. The first two categories are benign tissues, the third one is a risk factor for malignancy (we included this into one of the malignant classes to reduce false negatives) and the last one is malignant. Most cancer regions in the PZ tend to have a dark appearance [12, 16]. Moreover, several studies suggested that prostate cancer tissue tends to appear darker on a T2-weighted MRI image [47, 48, 285]. In fact, radiologists also tend to use darker regions to identify abnormality within the PZ [49]. Since most malignant regions contain lower intensities, cancerous regions could be detected within the prostate by taking the segmented regions that correspond to the first two lowest intensity fuzzy c-means clusters (indicated by the superscript '*low*' in Figure 4.7) in G_1, F_1 and M_1. However, since malignant regions in F_2 are represented by higher average intensity values, we take segmented regions which correspond to the two highest intensity fuzzy c-means clusters. This process can be represented using the following equation

$$O = G_1^{low} \cap F_1^{low} \cap F_2^{high} \cap M_1^{low} \tag{4.14}$$

where '*low*' and '*high*' are low and high intensity represented in the segmented regions within the prostate and O represents the overlapping region from all four binary segmentations. Figure 4.7 shows an example of this process.

After selecting the regions of interest (segmented areas which are under the approximate PZ's boundary (green line in Figure 4.3)), we combine all binary segmentations and find its overlapping region as shown in Figure 4.7. Finally, we perform erosion to remove noisy pixels which will be explained in the next subsection. Note that segmented areas above the green line were ignored in this study because we are only interested in detection within the PZ.

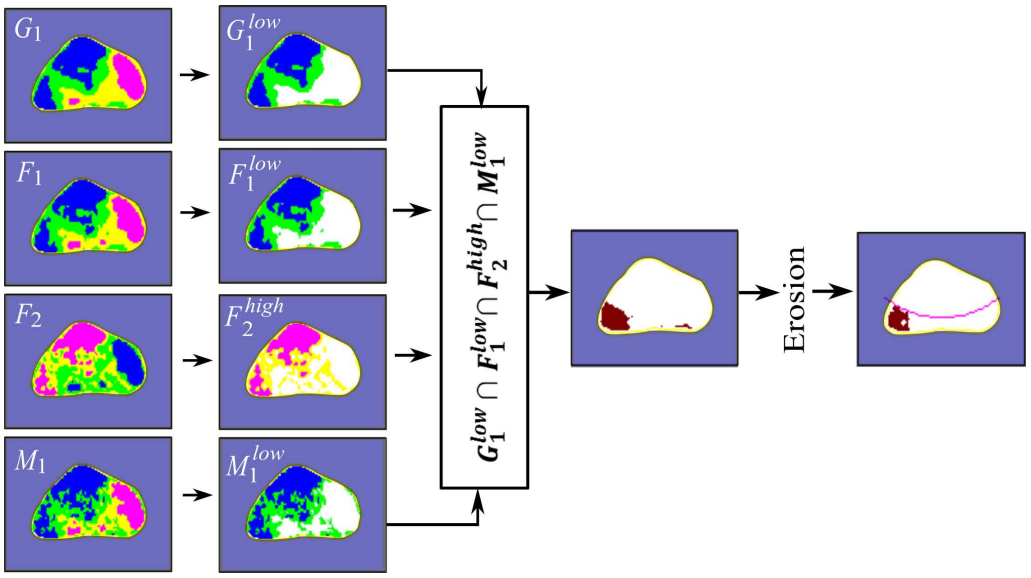

FIGURE 4.7: After performing FCM clustering on every feature, we only take segmented regions which correspond to the first two lowest intensity FCM in G_1, F_1 and M_1, and segmented regions which correspond to the two highest intensity clusters in F_2. Note that only segmented areas which are under the green line in Figure 4.3 (the estimation of the PZ's boundary) will be taken into account. Segmented areas above the green line were removed.

4.4.5 Post-processing

The post-processing step in this method involves erosion which is one of the basic operators in the area of mathematical morphology. The erosion operator aims to reduce the boundaries of regions of foreground pixels using the following rule '*the value of the output pixel is the minimum value of all the pixels in the input pixel's neighborhood. In a binary image, if any of the pixels is set to 0, the output pixel is set to 0'* [286]. By performing erosion on the binary segmentation, we can reduce the number of false positives. The number of pixels removed from the objects in an image depends on the size and shape of the structuring element used. Figure 4.8 shows how the erosion operator works on a binary image.

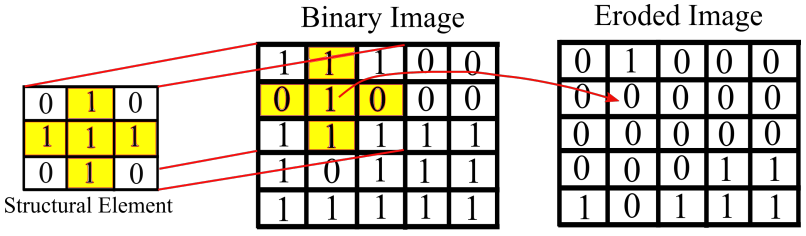

FIGURE 4.8: Using the erosion operator of 'disk' shape and size=1 on a binary image

The operator has two parameters: (a) the types of structuring element (e.g. disk, diamond, circle, flat, etc.) and (b) the size of the structuring element (the number of adjacent pixels from the pixel of interest or central pixel). In the proposed method we used a 'disk' shaped structuring element with size either 1 or 2. The 'disk' shape was chosen due to regions with ellipse-shape have higher change of being malignant [287, 288]. The size selection of the structuring element depends on the size of the segmented region within the peripheral zone. If the size of the segmented region within the peripheral zone covers $\geq 20\%$ of the size of the peripheral zone, the size of the structuring element is 2, otherwise 1. This ensures that every segmented region is not over eroded or under eroded during the process.

4.5 Database Descriptions

Data from 37 patients (range: 40-74 years) with biopsy-proven prostate cancer were included in this study. All patients underwent T2-W MR imaging at the Department of Radiology at the Norfolk and Norwich University Hospital, Norwich, UK. MR acquisitions were performed prior to radical prostatectomy. All patients gave their written consent to participate in this study which was approved by the institutional review board. All images were obtained on a 1.5 Tesla magnet (Sigma, GE Medical Systems, Milwaukee, USA) using a phased array pelvic coil, with a 24×24 cm field of view, 512×512 matrix, $3mm$ slice thickness, and $0.5mm$ inter-slice gap. Each patient has 5 to 12 slices. However, since our current study is focusing only within the PZ, slices with no visible PZ (the whole prostate gland is covered by the CZ) were excluded in this study (e.g. see Figure 4.9). All images were manually annotated by an expert radiologist (and further validated/confirmed by two independent radiologists) with more than 10 years experience in diagnosing prostate cancer in MRI. In total our database contains 275 slices taken from an axial direction (135 malignant and 140 normal slices). Each slice contains the annotations of prostate gland, central zone and cancerous regions (if present).

4.6 Experimental Results

Data was analysed and classified as to whether the prostate contains cancer. The detection of cancer occurs when there are any retained segmented regions ($G_1^{low} \cap$

$F_1^{low} \cap F_2^{high} \cap M_1^{low}$) within the peripheral zone. Subsequently, we compared the result with the ground truth whether the prostate contains cancer regions or not. We use several quantitative measures to evaluate the results such as sensitivity (Sen), specificity (Spe) and classification accuracy (CA). Each of these metrics can be calculated using the following equations

$$Sen = \frac{TP}{TP + FN} \tag{4.15}$$

$$Spe = \frac{TN}{TN + FP} \tag{4.16}$$

$$CA = \frac{TP + TN}{TN + TP + FP + FN} \tag{4.17}$$

where TP and FP denote the number of true positives and false positives, respectively. Similarly, TN and FN indicate the numbers of true negatives and false negatives. Accuracy means the number of correctly classified slices (or pixels in voxel based classification) out of the total number of slices. Sensitivity measures the proportion of actual positives which are correctly identified (in this case the percentage of malignant slices which are correctly identified) whereas specificity measures the proportion of actual negatives which are correctly identified (in this study the percentage of normal slices which are correctly identified). The proposed method achieved CA=86% (237 samples are classified correctly) and 38 samples data are misclassified with 7% (20 samples) false negative and 6% (18 samples) false positive results. In addition, the method produced Sen=87% and Spe=86%. Erosion with flexible size of structuring element and regions intersection ($G_1^{low} \cap F_1^{low} \cap F_2^{high} \cap M_1^{low}$) reduce the number of false positive and false negative results by $\approx 20\%$.

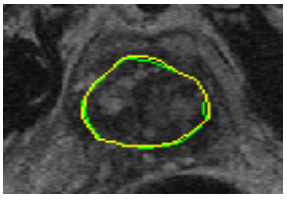

FIGURE 4.9: The whole of the prostate gland is fully covered by the CZ. All cases like this were excluded in our study because our current focus is within the PZ.

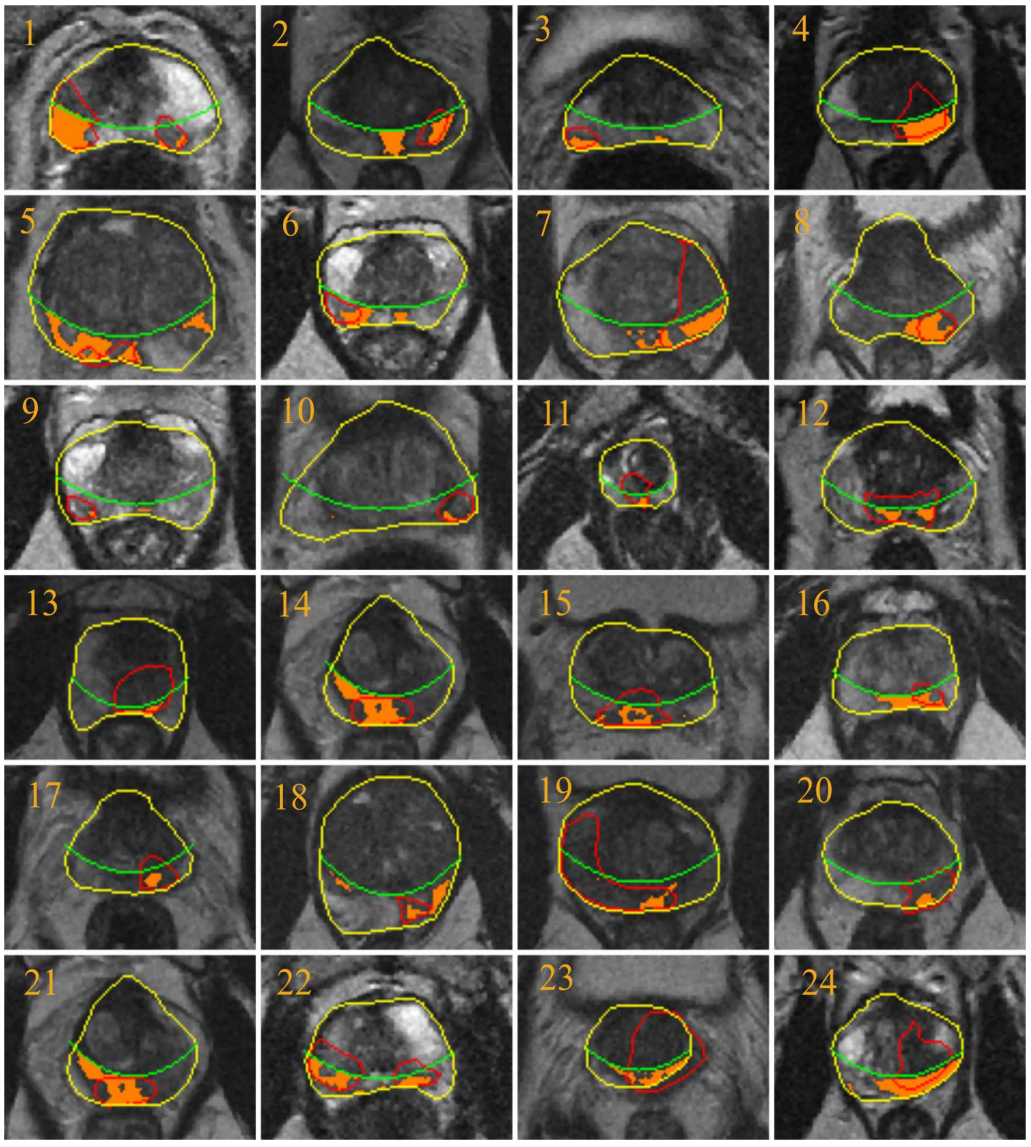

FIGURE 4.10: Malignant slices: prostate capsules are delineated in yellow and central zones and tumors are in green and red, respectively. The detected regions are indicated as the highlighted regions.

4.6.1 Correct Detection (Classification)

Figure 4.10 presents several examples of correct detection/classification. Correct detection means an image (MRI slice) is classified correctly (malignant or normal) regardless of the location of tumor within the PZ. The segmentation results are divided into three different categories; small malignant region, large malignant region and obscure malignant region. For the first category, the proposed method shows its sensitivity dealing with small malignant regions within the PZ as shown in image 3, 9, 10, 16 and 18. In those images, the proposed method managed to segment malignant regions correctly (in red line) despite their small sizes. In the second category, we

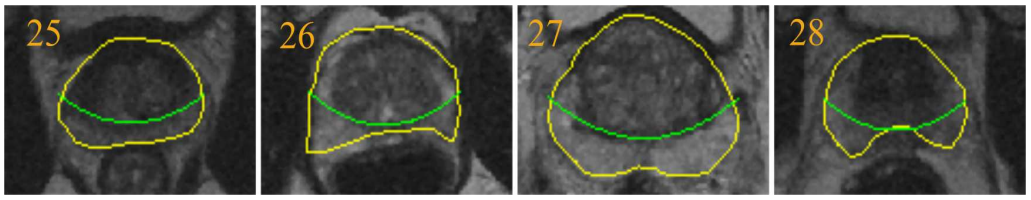

FIGURE 4.11: Normal slices: prostate capsules are delineated in yellow and central zones in green. The lack of segmented regions on the images indicates no cancer regions present.

show results in detecting and localising malignant region in larger areas. This can be seen in images 1, 4, 7, 13, 14, 15, 17, 19 and 20 where cancers are spread quite substantially within the PZ and some within the CZ. Results in Figure 4.10 show that these regions were segmented within the expert radiologist's ground truth. Finally (third category), we show results when malignant regions are obscure within the PZ. In image 2, there are three dark regions (left, middle and right) within the PZ and visually it is very difficult to identify which one of those regions is cancerous. As a result, although the proposed method managed to segment the malignant region, there is one false positive region in the middle of the PZ. On the other hand, in image 8 we can visually see that there is no sign of irregularity (the whole PZ looks uniform), which makes the abnormal regions obscured. Other examples of the experimental results can be seen in image 5, 6, 11, 12, 21, 22, 23 and 24. Figure 4.11 shows examples of experimental results in normal slices. The PZs in image 25 to 28 show no sign (or small signs) of irregularity which made it easier to identify normal slices.

4.6.2 Correct Detection with Incorrect Localisation

This section presents two examples of results where overall classification is correct but the tumor location is incorrect. For localisation, we compare the position of the segmented region based on $(G_1^{low} \cap F_1^{low} \cap F_2^{high} \cap M_1^{low}) \subseteq M_r$, where M_r is a cancerous region within the PZ. In our evaluation, incorrect localisation is when

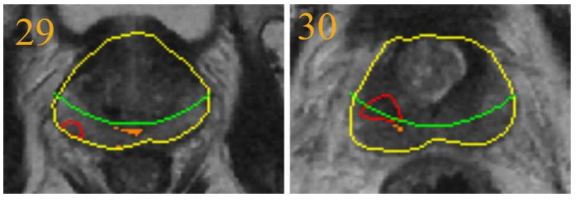

FIGURE 4.12: In both slices (image 29 and 30) the segmentation results show correct overall classification (true positive) but incorrect localisation in comparison to the location of the ground truth.

the area of the segmented region is $< 50\%$ within the cancerous region delineated by an expert radiologist. On the other hand, correct localisation means $\geq 50\%$ of the area of the segmented region is within the annotated malignant region. Our method produced 81% (109 slices true positives) correct localisation with respect to the number of malignant slices (135 samples) which means 6% (8 slices) of malignant cases were classified correctly but tumors were localised incorrectly and the other 18 slices are false negatives. This may have been caused when normal regions have dark or very similar appearance to cancerous regions (low intensity) in the PZ. According to [257] low signal intensity may be seen in the PZ on T2-weighted when blood products may persist after prostate biopsy. Moreover, when the location of the tumor is outside our PZ model (area under the green line in Figure 4.3). Figure 4.12 shows examples from our experimental results for correct classification but incorrect localisation.

4.6.3 False Positives and Negatives

Figure 4.13 shows four examples of false positive results from four different prostates. In images 31, 32, 33 and 34, there are clearly dark regions (higher probability of cancer) within the PZ which leads to false positive results. Based on the results in Figure 4.13, we can visually see that irregularity can occur in some normal slices which makes it hard to differentiate between malignant and normal regions. On the other hand, Figure 4.14 shows examples of false negative results from four different prostates. The malignant regions show obscure irregularity which lead to false negatives.

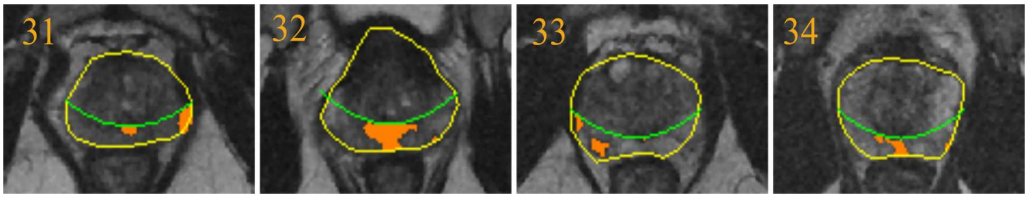

FIGURE 4.13: False positive results from four different prostates.

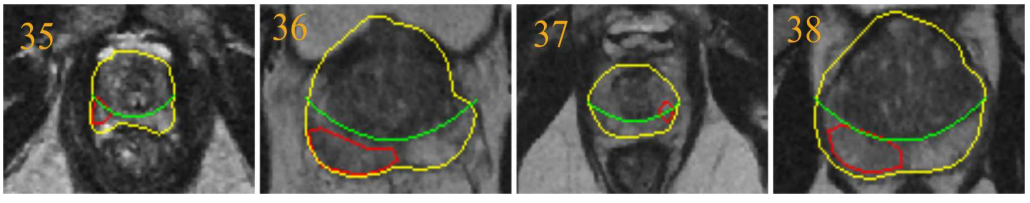

FIGURE 4.14: False negative results from four different prostates. Tumor regions are delineated in red.

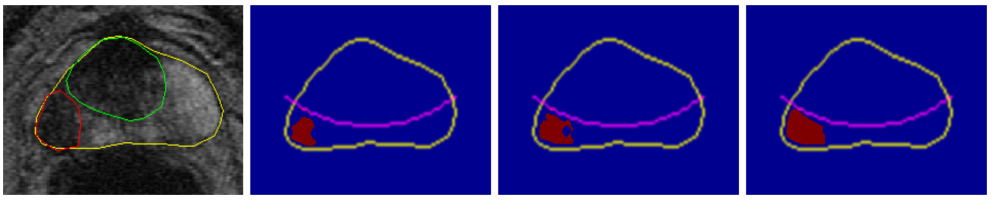

FIGURE 4.15: Large malignant region. From left to right, $\sigma = 3, 5, 7$ with kernel size 9×9, 15×15 and 21×21, respectively. The number of classified pixels are 189, 239 and 343.

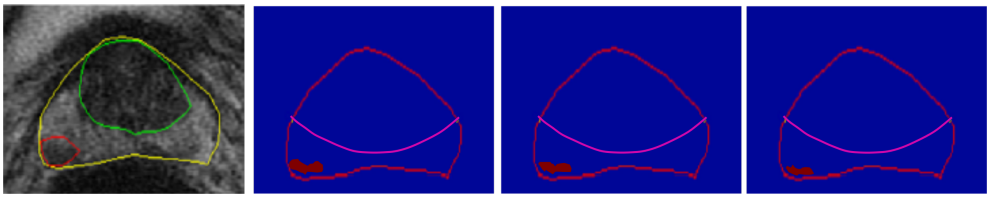

FIGURE 4.16: Small malignant region. From left to right, $\sigma = 3, 5, 7$ with kernel size 9×9, 15×15 and 21×21, respectively. The number of classified pixels are 59, 55 and 46.

4.6.4 Parameters Justification

Figure 4.17 shows the justification of our selected parameters. Based on the varying $\sigma = 2, 3, 5, 7$ and 9, the following kernel sizes $7 \times 7, 9 \times 9, 15 \times 15, 21 \times 21$ and 27×27 are applied, respectively. The results show that better sensitivity (above 80%) is achieved when $2 \leq \sigma \leq 5$. The sensitivity of the proposed method decreases when $\sigma > 5$ due to the level of smoothing applied to images. For instance, higher value of σ would affect (e.g. over-smoothed) the appearance of small malignant regions, hence decreases the sensitivity. On the other hand, the method achieved its highest specificity when $\sigma = 5$. Using $\sigma \leq 5$ still gives similar sensitivity to the other methods in the literature. Although varying the σ and kernel size did not change the results significantly, we have shown the quantitative results for the justification of our selected parameters. However, varying these parameters certainly affect the areas of segmented regions. Figure 4.15 shows close up examples of segmented region using different values of σ and kernel sizes together with their quantitative segmentation results (number of classified pixels). It shows that larger σ values benefit larger malignant region but reduce the segmented area for a smaller malignant region as shown in Figure 4.16. Too large σ value (e.g. 50) would over smooth the image and the abnormality might disappear which lead to false negatives.

Figure 4.18 shows results using different ϵ values. Our experimental results show that $\epsilon = 0.75$ produced balanced results in terms of accuracy, sensitivity and specificity. A

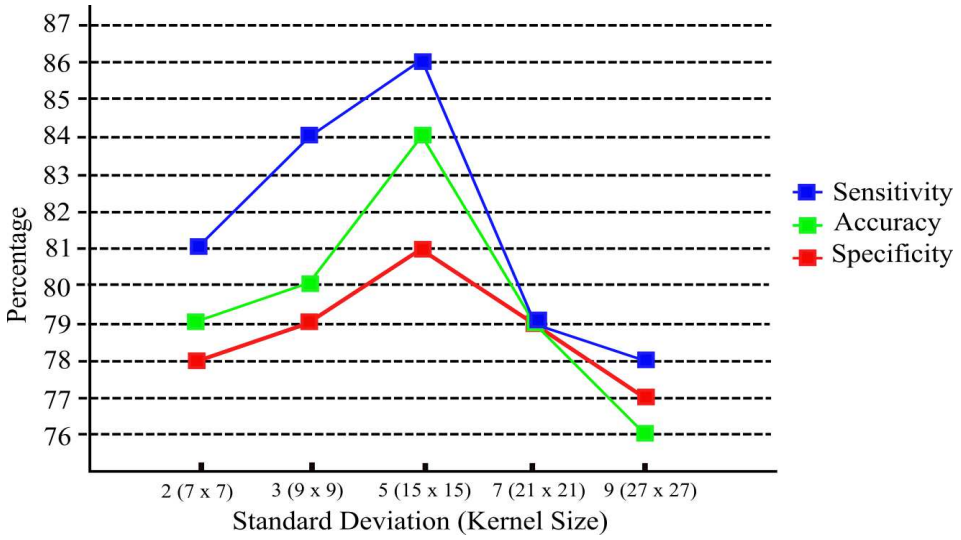

FIGURE 4.17: Sensitivity, specificity and accuracy using different values of σ and ks.

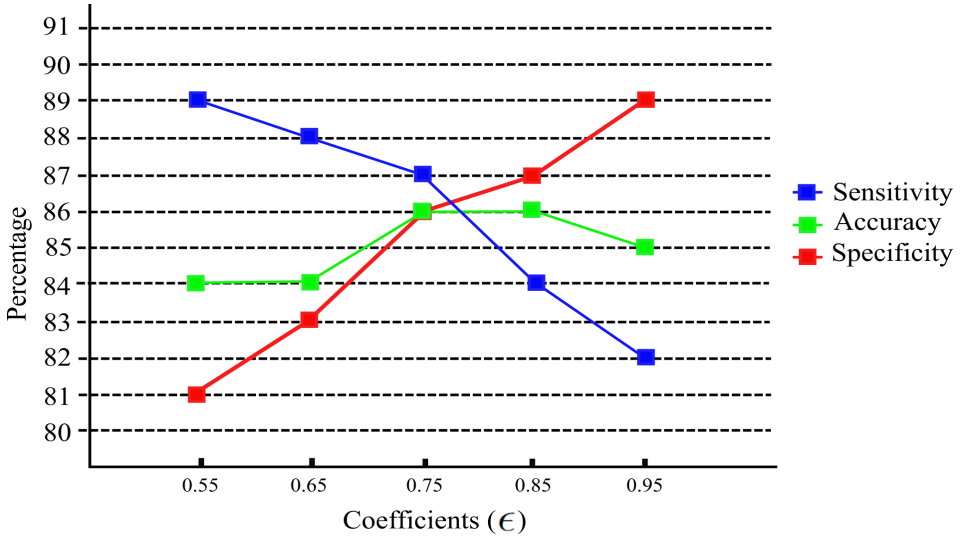

FIGURE 4.18: Coefficient $\epsilon = 0.75$ produce balanced results, with higher specificity for larger ϵ values and higher sensitivity for lower ϵ values.

larger ϵ value (e.g. 0.95) reduces the area under the curve (green line in Figure 4.3), hence most cancerous tissues were missed which increased specificity but reduced the algorithm's sensitivity. On the other hand, a smaller ϵ value (e.g. 0.55) increases the area under the curve (green line in Figure 4.3). This increased the algorithm's sensitivity (and false positives) because the larger area leads to a higher chance of cancerous tissues being detected.

4.7 Discussion

The proposed method produced similar accuracy, sensitivity and specificity to the state of art in the literature particularly in single modality T2-Weighted MRI. However, due to various factors a direct comparison is less appropriate but will be discussed in Chapter 7. In contrast to the earlier methods, our method is different in the sense that:

1. The proposed method does not need a classifier model, hence no training phase is needed to be able to discriminate malignant and benign tissues. In contrast to most of the methods [21, 192, 206, 227, 229, 231] in the literature they used machine learning techniques to build predictive models and subsequently used the models to make prediction (e.g. benign or malignant) in testing unknown cases.

2. We only used a single modality for abnormality detection which is T2-Weighted MRI whereas majority or the studies in the literature combined different modalities to extract more different features. The methods in [54] used multimodality such as diffusion MRI and MR Spectroscopy. Similarly, the method proposed in [21] used a multiparametric MR of T1- and T2-weighted imaging. Engelbrecht *et al.* [54] suggests that various techniques such as dynamic contrast material enhanced MR imaging, diffusion-weighted imaging, and MR spectroscopy have the potential to improve the detection of prostate cancer. On the other hand [289] combined the use of diffusion-weighted MRI and 1H MR Spectroscopy to get better results in discriminating malignant and normal tissues.

3. The method in [287] used additional clinical knowledge (e.g. location and shape of the region) to discriminate cancer regions in addition to image features while the proposed method is entirely relaying on intensity-based image features to achieve similar results.

4. Our method used a small number of image features to discriminate malignant and benign regions and produced similar results to the state of art in the literature whereas the methods in [287, 290] used more features.

5. The methods in [243, 291–293] used various perfusion parameters on a single modality while our method is purely based on image features but still managed to achieve similar results.

In our study, one obvious drawback of the proposed method is the risk of classifying correctly with incorrect localisation of the tumor, which could be problematic from a clinical point of view. Secondly, in some cases when the prostate's peripheral zone is almost non-existent, the proposed method is more likely to produce false positives. This is due to the intensities being very similar between the central gland and the malignant region [257]. Finally, if the prostate's shape does not conform to the shape of our prostate model a smaller area of the PZ will be analysed which may increase the chance of malignant regions being missed. Therefore, in order to accommodate these limitations for future work we are planning to use a multiparametric approach (e.g. T2-W+DWI+DCE) instead of stand alone T2-W MRI. This means more image features can be extracted which could help to distinguish malignant and benign tissues. In addition, we intend to cover the whole prostate gland instead of only the PZ.

4.8 Summary

In summary, the proposed method uses the following image features: a) G_1 and M_1 are extracted from Gaussian and median filters, and b) F_1 and F_2 are the sum vectors of the probability images (G_2 and M_2) which were extracted from G_1 and M_1. The classification of malignant prostate and localisation are defined based on $G_1^{low} \cap M_1^{low} \cap F_1^{low} \cap F_2^{high}$ and $(G_1^{low} \cap M_1^{low} \cap F_1^{low} \cap F_2^{high}) \subseteq M_r$, respectively within the peripheral zone. Erosion and regions intersections are applied to reduce false positives and false negatives.

The proposed method delineates regions which have the highest probability to be malignant (see results in Figure 4.10), hence help radiologists to perform targeted biopsies and potentially improve the accuracy of prostate cancer diagnosis [294]. We have presented a novel method of prostate cancer detection and localisation within the PZ and successfully applied it on 37 patients covering 275 MRI images. Gaussian and median filters together show promising potential to be effective texture descriptors to identify cancer regions within the peripheral region. The other two features are image magnitudes computed from the probability image and median and Gaussian filtered images. The proposed idea, which is based on regions intersection and flexible size of erosion's structuring element, suggest a good potential to reduce false positives and false negatives in the proposed method.

In the next chapter, another CAD method will be proposed using a large number of texture features extracted and various types of machine learning techniques will be employed. Chapter 5 will cover 215 texture descriptors and each of them will be evaluated individually and ranked to find out which of them has the best ability to differentiate benign and malignant regions. Furthermore, the effects of using different window sizes also will be assessed the texture descriptors' performance. On the other hand, 11 machine learning techniques will be employed and quantitatively compared to find out which technique has the best predictive model.

Chapter 5

Multifeature with Machine Learning Algorithms in Prostate Cancer Detection

In this chapter, we propose a CAD-PC using various features mainly motivated by statistical, psychological and image and signal processing points of view. Eleven different classifiers were employed to achieve the best possible results. For this purpose, 215 texture descriptors were extracted divided into six categories namely first-order, second-order, Tamura's features, gradient-based features and filter bank. The proposed method was tested based on 418 T2-weighted MR images taken from 45 patients and evaluated using 9-fold cross validation with five patients in each fold. The results demonstrated comparable results with existing CAD systems using multiparametric MRI. We achieved area under the receiver operating curve (A_z) values equal to $90\% \pm 7.6\%$, $89.5\% \pm 8.9\%$, $87.9\% \pm 9.3\%$ *and* $87.4\% \pm 9.2\%$ *for Bayesian Networks, ADTree, Random Forest and Multilayer perceptron classifiers, respectively, while a meta-voting classifier using average probability as a combination rule achieved* $92.7\% \pm 7.4\%$. *Results are comparable to the state-of-the-art models both CAD-PC based on single and multi-modality.*

5.1 Introduction

Many studies have reported the limitations of computer-aided diagnosis systems using single modality T2-weighted MRI, which can include weak texture descriptors and an

excessive amount of noise. Researchers have used multiparametric MRI to improve the performances of their methods. However, CAD methods based on multiparametric MRI have their own limitations as discussed in Chapters 1 and 3. Therefore, the ultimate goal here is to develop a CAD-PC demonstrating that a CAD system based on a single modality of T2-W MRI is capable of achieving similar performances to those based on multiparametric MRI for prostate cancer detection. Although previous studies conducted by Lv *et al.* [211] and Lopes *et al.* [208] have shown CAD systems based on single modality achieved A_z>90% using fractal features, their studies need to be further evaluated due to small numbers of datasets 130 and 104 region of interests (ROI), respectively. In fact, Lv *et al.* [211] did not perform cross validation to further evaluate the results of their method. The novel contributions of this chapter are the following:

1. The proposed method incorporates a large number (215) of different texture descriptors from T2-W MRI. This means our study investigates a number of novel feature options that have not previously been applied in CAD-PC. To the authors' knowledge the previously used largest number of 2D texture descriptors in the literature using only T2-W MRI was in a study conducted by Viswanath *et al.* [295] (110 texture descriptors) and 83 texture descriptors by Tiwari *et al.* [65]. Using multiparametric MRI Niaf *et al.* [231] extracted 140 texture descriptors from T2-W MRI, diffusion imaging (DWI), and dynamic contrast enhanced (DCE).

2. We extensively compare 9 classifiers' performances with two additional combined classifiers (11 classifiers in total). Again, to our knowledge the largest number of classifiers used in the literature in CAD-PC is in a study conducted by Niaf *et al.* [231] (4 classifiers), and Litjens *et al.* [46] and Ozer *et al.* [198] employed 3 classifiers.

3. Since this study involved a large number of texture descriptors, we evaluated all features individually and combined them to improve performances of the proposed method. By evaluating them, we are able to determine which features individually give the best performances on each classifier.

4. Finally we investigate the effect of different window size on the performance of the proposed method as well as the performances of the features individually. To our knowledge, none of the current CAD-PC in the literature have reported

quantitatively the effect of window sizes on performance. Many studies [65, 231, 233, 263, 296] selected window size without a qualitative justification.

These contributions are expected to be beneficial to the research community as they provide state-of-the-art CAD-PC using single modality T2-W MRI. Moreover, this study provides general guidelines on selecting window size, classifier as well as the set of features in CAD-PC development.

5.2 A Review of Machine Learning Algorithms

In this section we briefly review 11 machine learning algorithms (or classifiers). Many machine learning techniques have been developed and described in the literature over the last decades. A full review of machine learning algorithms in pattern recognition can be found in [297]. It is difficult to categorise each method, as many method have similarities in terms of implementation, structures, etc. However, in general these algorithms can be divided into 11 categories [298]: deep learning, ensemble, neural network, regularisation, rule system, regression, bayesian, decision tree, dimensionality reduction, instance based and clustering. In the area of pattern recognition of prostate cancer imaging, these algorithms are known as classifiers which are categorised into the following methods: rule-based, clustering, linear model, non-linear model, probabilistic, ensemble, kernel-based, neural network and graphical model based classifiers. In the study of CAD-PC, kernal-based and ensemble methods have gained the highest popularity among researchers in the last decade. To be more precise Support Machine Vectors (SVM) and Random Forest (RF) have showed high discrimination capability between benign and malignant tissues. Therefore in this chapter, these two classifiers will be employed to build statistical predictive models. However, SVM is very sensitive in relation to its parameter settings whereas RF is very robust and less dependent on its parameters.

The CAD-PC studies in [21, 45, 65, 194, 196–198, 208, 226–228, 233, 291, 299, 300] built SVM models for classification purpose. SVM is a supervised learning method which aims to find the best linear hyperplane which separates two classes by maximising the margin between them [301, 302]. However, many classification tasks in real world cases are non-linear and linear SVM is not able to handle such datasets. For a non-linear boundary between two classes a non-linear mapping function will be used to transform the data into a new feature space (by computing the feature map),

which means data will be projected into a higher-dimension space (e.g. from 2D to 3D) so a hyperplane which can separate those classes can be found. This is based on the assumption that data which is linearly nonseparable in the original feature space may be linearly separable in a higher-dimensional space. Nevertheless, computing the feature map can be computationally expensive and inefficient. These issues can be solved using 'kernel trick' by computing the dot product between two vectors without transforming these vectors into a higher-dimensional space. There are a few popular kernel functions which are often used in SVM such as Polynomial, Gaussian and Radial Basis Function (RBF) and Sigmoid kernel.

In contrast, the CAD-PC studies in [46, 65, 229, 238] employed ensemble learning classifiers. The RF classifier was used in [46, 65, 229, 238] is among the most popular classifier in the study of CAD-PC. The RF classifier builds a large number of learning models using decision trees from various sub-samples of the training dataset to increase the classification accuracy [303]. By employing different learning models to create a class boundary, a model with low variance can be created by averaging noise and unbiased models which is the main idea of a bagging technique [304]. The training data will be divided into many sub-samples (also known as forests) randomly (e.g. 100 forests) and a decision tree will be created in each of the sub-samples (hence, 100 forests will have 100 decision trees). Each of these decision trees will be used to make a prediction in a test sample and the final decision will be determined via the majority vote among the decision trees. For example if 60 decision trees predicted the sample is 'benign' and 40 decision trees predicted the sample is 'malignant' then the decision result will be 'benign'.

A couple of decision tree based classifiers will be employed in this study namely C4.5 (also known as J48) [305] and alternating decision tree (ADTree) [306]. The C4.5 classifier is based on the top-down, recursive and conquer strategy. In this method, the best attribute is first selected to split the root node. The best attribute is the one which has the purest nodes which means all the instances at that node have the same class label and no need to further split. This can be determined by measuring the amount of 'information entropy' for each attribute (the highest amount of 'information entropy' will be the best attribute) [305]. Once the best attribute is selected, a branch for each possible attribute value (e.g. let say we have five branches) is created. The first branch is then split based on the attribute which has the highest value of 'information entropy' (excluded the first attribute selected) which will split instances into subsets. Similarly, the second branch will be split based on the attribute which

has the second highest value of 'information entropy'. The same procedure will be applied to split the third until fifth branch. Repeat this process recursively for each branch until all instances have the same class label (no need to further split).

The ADTree is another classifier which uses the concept to map the classification procedure in a form of a decision tree but using 'boosting' approach to improve the classification accuracy [306]. The 'boosting approach' [307] is a learning algorithm which iteratively builds a weak classifier based on the accuracy of the previously built weak classifier. This makes the later weak classifier more tolerant to the misclassified data in the previous iteration. In a conventional decision tree, the root node represents the best attribute but in ADTree it is a prediction node simply contain a numeric score (prediction value) [306]. The prediction value is a weight computed based on the error rate produced by the previous week classifier. The next layer of the node are called decision nodes which specify the predictive condition (contains attributes). The tree's structure will be alternating between prediction nodes and decision nodes. An instance is classified by following all possible paths for which all decision nodes are true from the root node to the leaves through the whole tree. The classification score (also represents the confidence level) is computed by summing up the prediction values along the corresponding paths [308].

The Naïve Bayes (NB) [309] classifier is based on the Bayes' theorem [310] where the algorithm assumes that all the relationships among attributes are conditionally independent of each other. It does not have iterative parameters which makes it easy and very fast, hence it is very popular in applications with very large datasets, and also making it quite popular in several CAD-PC studies such as [217, 231, 236, 311]. In this study, the basic idea behind the Bayes' theorem is to calculate the probability of an instance classified as being malignant or benign based on the observed data from the training sample. The probability of an instance being classified as malignant is evaluated by computing the probability of its attribute value occurring in the class 'malignant' of the training data [309]. However, since the NB classifier assumes all attributes are conditionally independent [309], all probabilities of attribute values in an instance have to be multiplied to get a single probability value (which is the product of probabilities). The value of the final probability represents the degree of likeliness of being classified into a particular class and the classification results will be based on the highest probability. Therefore for an instance (or test data) with multi-attribute, to compute its probability of being in a particular class, it can be

computed using the following equation

$$P(class|A) = P(a_1|class) \times P(a_2|class) \times .....P(a_f|class) \times P(class) \quad (5.1)$$

where $A = a_1, a_2, a_3....a_f$ consists of attribute values and $P(a_n|class)$ is the probability of the predictor in a given class and $P(class)$ is the probability of a given class.

In contrast, Bayesian Network (BNet) [312] does not have assumptions like the NB (attributes are conditionally independent of each other) but computes the probability dependencies between attributes. The map/graph structures (each node is connected with the other nodes) are used to represent knowledge about an uncertain domain. Each node in the graph represents an attribute which contains a random variable while the arch between the nodes represents the probabilistic dependencies among the corresponding random variables. This means the probabilistic relationship among two random variables (of attributes) is computed by the BNet algorithm based on the Bayes' rule [310]. The attributes in the training data are first mapped in a form of graph topology based on the probability dependency values among the attributes. This can be done by computing the probability of the random variable occurring in the training data. For a random variable (A) which is dependent on another variable (B), the algorithm computes the probability of A given that B has occurred (Bayes' rule [310]). In the testing phase, to determine whether an instance belongs to the 'malignant' or 'benign' class, the BNet computes the probability of being in the class 'malignant' given the attributes and the joint probability of the attributes [312]. The summation of these two probabilities values will be the confidence value of whether an instance belongs to a particular class or not.

The k-Nearest Neighbour (k-NN) is one of the simplest supervised learning algorithms and is very popular in image classification. Nevertheless, in the study of CAD-PC, this method is less popular as it did not not produce as good results as the SVM and RF classifiers. In fact, studies in [228, 231, 236] used k-NN just to make a comparison with different machine learning algorithms. The prediction of an instance belonging to a particular class is determined by the closest neighbour(s) in a feature space within k neighbours, where k is the number of neighbouring instances (from the training data) set by the user. For example, a test data (t^d) with three attributes will be mapped into three dimensional feature space and (let k=5) the algorithm will consider five closest neighbours (using Euclidean distance). Subsequently, k-NN will check whether the instance is surrounded by neighbouring instances of 'malignant' or

'benign' class. A majority vote approach will be used for this purpose. For example if three of the five neighbouring instances labeled as 'malignant' classes then t^d will be classified as belonging to 'malignant' class. Another classifier employed in this method is Simple Logistic (SL) [313], which is a popular classifier for building linear logistic regression models. The SL classifier contains two main approaches: firstly a boosting algorithm called '*LogitBoost*' [314] is used to improve the classification accuracy and secondly simple linear regression (SLR) functions were used as base learners [314]. This means, for each iteration a base learner will be constructed and the least-squares fitting technique is used to fit the base learner (weak learner) into the logistic regression model. The base learner is created using SLR by finding the linear relationship between dependent and independent variables [313]. In this algorithm, the number of boosting iterations represents the number of base learners. The '*LogitBoost*' algorithm is one of the variations of boosting approach, however in comparison to the conventional boosting (e.g. majority vote, averaging, etc), at each iteration the 'response variables' is calculated to estimate the error of the current model (logistic regression model) on the training data and a weak learner is fitted (or added) to improve the current model [313]. The process is repeated until the value of least-squared error is minimised (hence a final model is achieved).

Multilayer perceptron (MLP) [315] is based on the neural network learning algorithm inspired by the biological neural network such as brain. The perceptron consists of weights (including bias), the summation processor and an activation function (e.g. threshold function). The weights are numeric values which will be adjusted accordingly by the algorithm to reduce the error rate and the summation processor is the sum of the inputs and bias whereas the activation function is the threshold function which will determine the classification result. For example, an instance with five feature values is fed into the perceptron and if the predicted outcome is the same as the expected outcome, then there is no need to change the weights [316]. Otherwise the weights need to be changed to reduce the error. The MLP algorithm is a solution of the single linear perceptron to solve non-linear separable data using the back propagation method. Indeed the difference is that the MLP contains L layers ($L > 1$) whereas single linear perceptron contains only one layer. Therefore, the inputs of L^{th} hidden layer is the outputs of the $L^{th} - 1$ layer and the outputs of the last hidden layer is the input of the output layer which will calculate the output error and result. The idea behind the back propagation method is that the weights of the previous layer which can be adjusted in reverse if the outcome is not the same with the expected

result. Studies which have employed the MLP classifier achieved high specificity but low sensitivity in [299].

Finally, the proposed method employed a traditional ensemble method which selects the two best (M-V(b2)) and three best (M-V(b2)) classifier (best area under the curve (AUC) values) based on the results produced after the training phase and use average probability as a combination rule [317]. This means the probability value (or confidence value) produced by each of the selected classifiers will be averaged and the mean of confidence values will be used to determine the final classification (e.g. if the mean of confidence value is more than 0.5 is considered malignant). To select the two best (or three best) classifiers, in the training phase, we performed 9-cross validation on the training data for all classifiers and select those have the highest two and three ROC values and combined them as an ensemble classifier.

5.3 Methodology

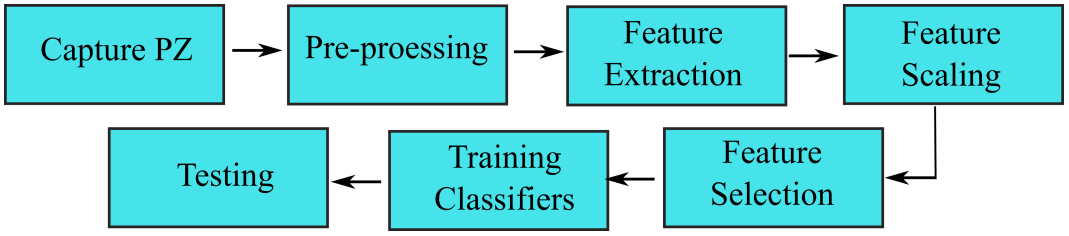

FIGURE 5.1: A general overview of the proposed method.

Figure 5.1 provides a general overview of our method. For every input image, we estimate the area of the prostate's PZ and extract 215 feature descriptors. We normalise each of the selected features. Feature selection was performed to eliminate irrelevant or redundant features and use them to train 11 classifiers. Finally in the testing phase, for every unseen pixel within the PZ the trained classifiers determined whether it belongs to the malignant or benign class.

5.3.1 Capturing the Peripheral Zone

Since segmenting the PZ manually is time consuming, we employed the method in Chapter 4 to segment the PZ, which uses a quadratic equation based on the central coordinates of the prostate gland, the left-most and right-most coordinates of the

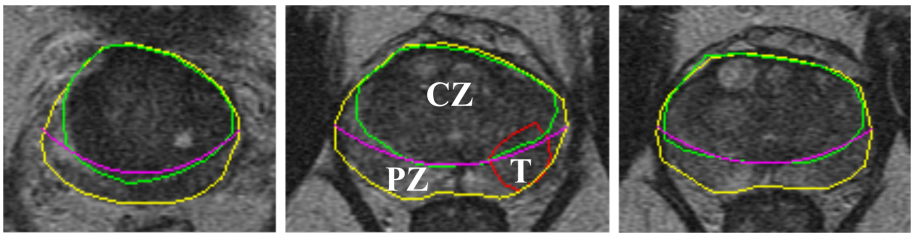

FIGURE 5.2: Example images of prostate MRI with the ground truth delineated by an expert radiologist and the estimated PZ region under the magenta line.

prostate gland boundary. Figure 5.2 shows example MRI images with the ground truth of prostate gland, central zone (CZ) and tumor (T) represented in red, yellow and green, respectively, while magenta line is the estimated boundary of PZ.

5.3.2 Pre-processing

The major problem in MRI image analysis is that intensities do not have a fixed tissue specific numeric meaning even within the same MRI protocol, the same body region, and the same scanner [189–191]. These problems are mainly caused by [189–192]: a) corruption by thermal noise due to receiver coils, b) intensity variations due to different scanning protocols and c) poor radio frequency coil uniformity. Intensity variations in MR data can significantly affect performances of many image processing techniques, hence, they need to be corrected [191]. Following the pre-processing procedure method described in [192, 204, 205], each image is median filtered to preserve edge boundaries. Subsequently, image intensities were normalised to zero mean unit variance and anisotropic diffusion filtering [205, 318] is applied to remove noise. The basic anisotropic diffusion equation presented in [318] is defined as

$$\frac{\partial I(x,y,t)}{\partial t} = div[f(\parallel \bigtriangledown I(x,y,t) \parallel).\bigtriangledown I(x,y,t)] \tag{5.2}$$

where t is the time parameter (number of iterations), $\bigtriangledown I(x,y,0)$ is the original image (the $t = 0$ indicates no iteration has been performed), $\bigtriangledown I(x,y,t)$ is the local image gradient at t^{th} iteration and $f(.)$ is the conductance (or flux) function. Perona and Malik [318] suggested two conductance functions:

$$f(x) = exp\left[-(\frac{x}{K})^2\right] \tag{5.3}$$

$$f(x) = \frac{1}{1 + (\frac{x}{K})^2} \tag{5.4}$$

Previously, Black *et al.* [319] introduced a more robust flux function which is able to preserve sharper edges defined as

$$f(x) = \begin{cases} 0.67[1 - (\frac{x}{K\sqrt{5}})^2] & ,x \ \leq k\sqrt{5} \\ 0 & , \text{otherwise} \end{cases} \tag{5.5}$$

The idea behind anisotropic diffusion is that it diffuses the image more in homogeneous areas, but not cross region boundaries and less near the inhomogeneous areas and edges. In contrast to the conventional smoothing techniques, the anisotropic diffusion has the knowledge of the region boundaries (using flux function) and thus preserves them [205, 318]. This reduces the noise of an image without blurring the edges because only pixels within the homogeneous region were smoothed. In comparison to isotropic diffusion techniques, the new value of the central pixel is computed by including neighbouring pixel values (although they come from different regions) in the calculation which means pixels located along the boundaries (or edges) are blurred. This will have similar effects in other techniques such as moving average filter. The following steps briefly describe the process of anisotropic diffusion filtering, using Figure 5.3 as pixels reference:

- Step 1: I is convolved with four gradient operators separately to find the gradient images (see [320] for the convolution masks used), resulting $\bigtriangledown I$.
- Step 2: The smoothed image at t^{th} iteration is computed as follow [321].

$$I_{t+1}(s) = I_t(s) + \frac{\lambda}{\eta_s} \sum_{p \in \eta_s} f_k(|\bigtriangledown I_{s,p}|)) \bigtriangledown I_{s,p} \tag{5.6}$$

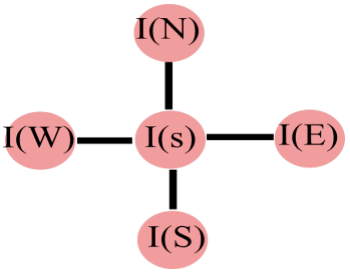

FIGURE 5.3: $I(s), I(N), I(S), I(W), I(E)$ represent the the central, north, south, left (west) and right (east) pixel, respectively.

where s denotes the pixel position in a 2-D grid (e.g. see Figure 5.3), K is the gradient threshold parameter and the constant λ controls the diffusion rate at any point in the image. On the other hand, η_s represents the spatial 4-pixel neighbourhood of pixel s (see Figure 5.3): $\eta_s = N, S, E, W$.

$$\bigtriangledown I_{s,p} = I_t(p) - I_t(s), p \in \eta_s = N, S, E, W \tag{5.7}$$

This method is chosen because it does not cause inter regional blurring [192, 204, 205]. However, [192, 204, 205] suggested that anisotropic diffusion filtering needs to be applied on the median-filtered and normalized images for better results. Therefore parameters for the anisotropic diffusion methods were based on the study conducted in [205]. The authors suggested the median filter with the smallest window size of 3×3 for the following parameters of anisotropic diffusion: K=the standard deviation of the image gradient, $\lambda = \frac{1}{7}$ (a small number is used to avoid destabilizing the diffusion process [267, 318]), the number of iteration=5 (suggested in [205]) and we used a *Tukey biweight norm* as the diffusion coefficient proposed by Black *et al.* [319]. This three-step pre-processing approach has the following advantages [192, 205] a) while suppressing the noise, it simultaneously preserves the edge boundaries b) it standardises image intensities for all patients avoiding dissimilar intensity values for the same tissue types c) it is a robust denoising method without blurring the tumor nodule edges. Other denoising techniques in the literature could also be investigated [266]. Figure 5.4 shows a visual comparison between mean, median and anisotropic diffusion filtering method. The resulting patch using mean filtering looks blurred whereas the texture in the patch using median filtering looked jagged. On the other hand, the texture using anisotropic diffusion looks preserved and smooth in comparison to the original image.

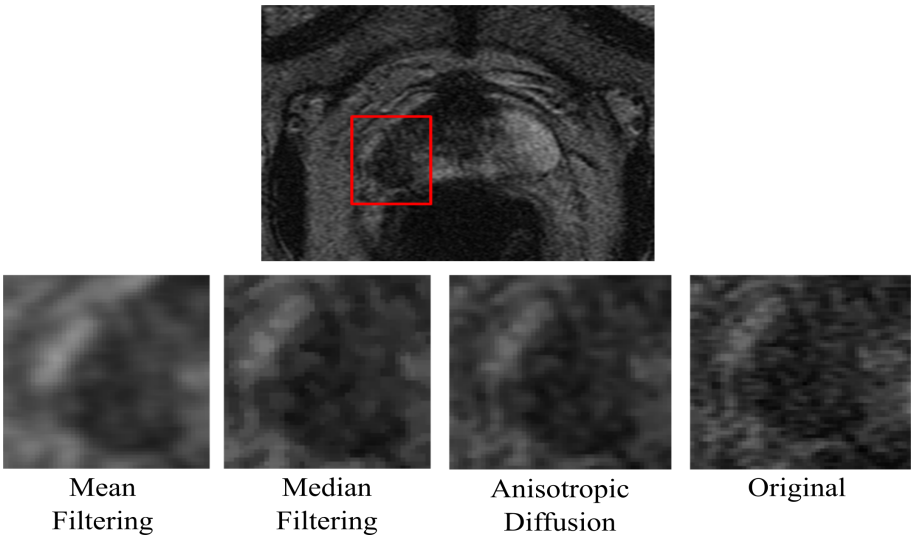

FIGURE 5.4: A visual comparison between filtered patch using mean, median and anisotropic diffusion filtering.

5.3.3 Feature Extraction

In this study, we extracted a set of 215 image features where the selection was mainly motivated by statistical, psychological and image and signal processing points of view. This means each pixel is represented in a 215 dimensional feature space. These features were selected based on the visual characteristics of malignant regions as indicated by expert pathologists as well as their efficiency in discriminating between malignant and benign regions [1, 174, 263]. However, although the bank of filters of Varma and Zisserman [1] and Tamura's texture features [174], have never been applied in CAD-PC, their studies [1, 168, 174] showed excellent results in texture classification. In this study, for first and second-order statistical features, Tamura texture features and grey-level percentile based features and estimated for each pixel for a local $n \times m$ window [231] where n and m are rows and columns, respectively. The features extracted in this study can be divided into the following categories:

First-order statistical features ***(F_1)***. These features mainly rely on image intensity as used in many pattern recognition algorithms. Studies in CAD-PC [198, 206, 231, 263, 296] have used intensity-based feature extensively in the last decade. From a clinical point of view, most malignant regions in the PZ tend to have a dark appearance (low intensity) [12, 16] (we are aware that in some cases low intensity does not represent malignancy (resulting in false positive detection) due to inflammation and post-biopsy scarring [49], therefore other texture descriptors such as a filter bank were used in this study) and radiologists tend to use darker regions as the basis of

their a *priori* knowledge to identify abnormality within the PZ [49]. Moreover, several studies have suggested that prostate malignant tissue tends to appear darker on a T2-W MRI image [47, 48, 285]. Therefore the selection of intensity-based features is appropriate in this study. Niaf *et al.* [231] used mean, median and standard deviation in their study. In addition, we extracted mean and median absolute deviation, skewness, kurtosis, the mean of correlation coefficients, local contrast, variance and local probability (11 features in total). A study in [322] indicated that many malignant regions are more likely to have low contrast value. On the other hand, in the previous chapter we have used probability images to quantify the likelihood of every pixel/voxel belonging to specific tissues (e.g. tumor region).

Second-order statistical features (F_2). The main motivation of using Haralick's features (Grey Level Co-occurrence Matrix (GLCM)) [108] is that these texture features characterized homogeneity, grey-level transitions, and anatomical structures of the image [62] as well as its simplicity, large range of potential features and its popularity (second after intensity-based features in MRI [62]). Julesz [169] states that first order statistics alone are not sufficient for humans to discriminate between two textures. Hence, in order for a CAD system to be able to discriminate textures Madabhushi *et al.* [296] considers both first-order and second-order statistical features. The GLCM is defined as the joint probability of occurrence of two grey level values at a given offset both in terms of distance and orientations [110]. We extracted all features originally suggested by Haralick *et al.* [108] and all features which were further suggested by Conners *et al.* [112]. To maximise the texture information captured from the co-occurrence matrix we considered four orientations ($\theta = 0°, 45°, 90°$ and $135°$) with distance d limited to 1. In addition, we calculate the mean, variance and standard deviation of four orientations for each of the features (154 features in total). From a clinical point of view, according to [323] malignant regions have a higher degree of uniformity in T2-W MRI. In fact, studies in [324, 325] suggested that the distribution of malignant foci detected by biopsies in the peripheral zone of the prostate is homogeneous. To capture these characteristics in features we use GLCM as we can calculate uniformity (also known as energy) and homogeneity.

Percentiles based features (F_3) [58] namely percentile 25% and percentile 75% to quantify the symmetry of the image (or a region of interest) intensity [83] (2 features in total). Vos *et al.* [300] and Niaf *et al.* [231] extracted similar features and found that many malignant regions have smaller values of percentile 75% (2 features in total). In our case we compute these features by replacing the central pixel with 25% and

75% percentile value of the grey levels within a $n \times m$ sliding window. These features represent the distribution of signal intensity within a specified window (e.g. 7×7 or 9×9). A smaller value of percentile 75% indicates that the central pixel within a specified window is surrounded by pixels with low intensities which increases the probability of being malignant (most tumors display low signal intensity within the PZ in T2-W MRI [50]).

Tamura texture features ***(F_4)***. In [174] six texture features corresponding to human visual perception were proposed: coarseness, contrast, directionality, line-likeness, regularity, and roughness. However, from experiments testing the significance of these features with respect to human perception, it was concluded that only the first three features are very important [174]. Therefore we only use the first three features in this study which are coarseness, contrast and directionality (3 features in total). Note that in this study we extracted the original (or standard) Tamura texture features. Since many malignant regions are more likely to have low contrast value [322], our initial hypothesis is that Tamura's contrast feature is more effective than the one extracted from the GLCM because Tamura contrast captures the variation of grey-level range and the polarisation of the distribution of black and white whereas GLCM contrast only captures the intensity variations within a $n \times m$ window [326].

Gradient features ***(F_5)***. There are many operators (e.g. Sobel filter, Kirsch filter, etc) that could be used to extract these features. In this study we only selected the most discriminative ones according to the results by Niaf *et al.* [231], namely image numerical gradient at 0° and 90° orientations and image magnitude. Secondly, using Sobel operators we extracted image gradient at 0°, 90° and diagonal orientations and image magnitude (7 features in total). According to [173], gradient operators perform well in characterising micro-textures as well as providing more consistent behavior as a descriptor of pathologies than co-occurrence matrices. In the previous chapter, we used image magnitude as one of our texture descriptors to segment malignant regions.

Filter bank features ***(F_6)***. From a clinical point of view, most malignant regions show textural distortions in T2-W MRI [262, 263]. Litjens *et al.* [46] captured these characteristics in features using a Gaussian texture bank. However the conventional Gaussian texture bank is a) more sensitive to rotation (hence, rotated versions of malignant textures would be classified as non-malignant unless those rotated versions were included in the training set) and b) it does not incorporate spots/bars and edges. Therefore, we employed a filter bank as proposed by Varma and Zisserman [1]

which is rotationally invariant and takes edges and spots/bars into account (38 features in total). The filter bank consist of an edge and a bar filter, at 6 orientations ($\theta = 0°, 30°, 60°, 90°, 120°, 150°$) and 3 scales ($(\sigma_x, \sigma_y) = (1,3), (2,6), (4,12)$), Gaussian and Laplacian of Gaussian filters both with σ =10 pixels (in total 38 responses). Viswanath *et al.* [263] extracted texture features using a bank of Gabor filters but the results from their study showed that Haralick's features were more discriminant in capturing malignant regions within the PZ and features extracted from Gabor filters work better in detecting malignant regions within the CZ.

Figure 5.5 shows eight examples of image features used in this study and Table 5.1 summarises the list of features used in this study which are divided into six categories.

5.3.4 Feature Scaling and Feature Selection

Since we have 215 texture descriptors, feature selection is necessary to a) reduce over-fitting when building a classifier model as less data means less chance of making decisions based on noise, b) possibly improve accuracy because only the most relevant attributes are selected to build a classifier model (a study conducted by Niaf

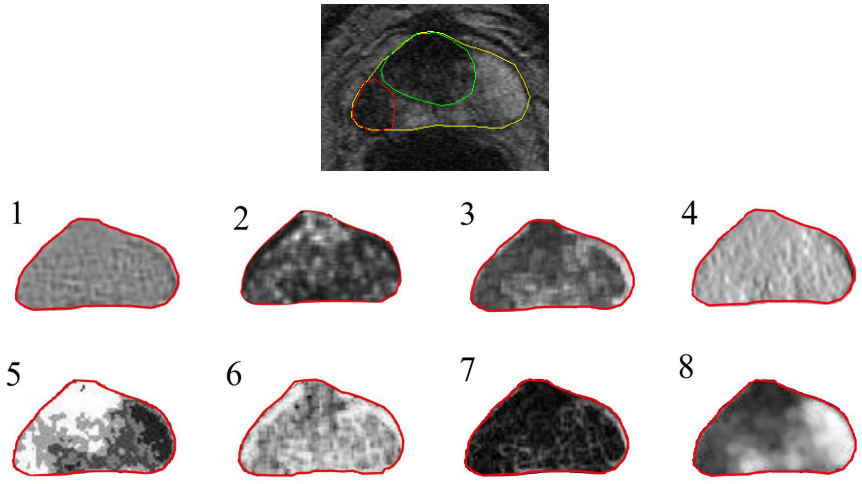

FIGURE 5.5: Image responses of some of the features used in this study. The top image is original image together with an annotation from a radiologist: red, yellow and green indicates tumour, prostate gland and central gland, respectively. The features are: Laplacian of Gaussian (image 1), GLCM- energy (image 2), local contrast (image 3), gradient of Sobel operator (image 4), probability image (image 5), GLCM- correlation (image 6), image magnitude of Sobel operator (image 7) and percentiles 75% (image 8).

TABLE 5.1: Summary of features used in this study.

Category	**Features**	**Total**
F_1:	Mean, median, standard deviation, mean and median of absolute deviation, skewness, kurtosis, mean of correlation coefficients, local contrast, variance and local probability	11
F_2:	GLCM features of Haralick *et al.* [53], Soh and Tsatsoulis [63] and Clausi [64] by taking 4 orientations and mean, variance and standard deviation of 4 orientations	154
F_3:	Grey-level percentile 25% and percentile 75%	2
F_4:	Tamura's textures features namely coarseness, contrast and directionality	3
F_5:	Image numerical gradient (0° and 90° orientations), image magnitude, using Sobel operator image gradient (0°, 90° and diagonal orientations) and image magnitude.	7
F_6:	Filter bank of Varma and Zisserman [30] which contains an edge and a bar filter, at 6 orientations and 3 scales, a Gaussian and Laplacian of Gaussian filters	38

et al. [231] demonstrated that feature selection significantly improves the discrimination performance between malignant and benign regions) and c) reduce training time because fewer features are used in making decisions. Many feature selection methods have been developed in the literature [225]. However, the main focus of this paper is not to select the best feature selection method but to build a robust CAD-PC using a single modality of T2-W MRI for prostate malignant diagnosis. Before feature selection is performed, we normalised each selected feature to avoid absolute values playing a role [235]. Following the suggestion in [235], each of the features was linearly scaled to the range [0,1] and the same was applied for the test data.

Subsequently, we employed the CfsSubsetEval [327] attribute evaluator and the GreedyStepwise search method in WEKA [2]. The CfsSubsetEval [327] method measures the value of a subset of features by considering each feature's predictive ability with the degree of redundancy between the other features within the subset, while the GreedyStepwise search method performs a greedy forward or backward search through the feature space [327]. Recently, Chen *et al.* [328] used the same method in classifying microcalcification clusters in mammograms.

5.3.5 Data Clasification

In this study we employed 11 different classifiers to achieve the best possible results using the WEKA data mining suite [2]. These classifiers were selected due to their robustness and popularity in CAD-PC [62]. Since our study evaluates a large number of classifiers, tuning parameters for each of the classifiers is time consuming and computationally expensive. Therefore, all parameters were left on default settings. The classifiers used in this study are presented in Table 5.2 (which includes their abbreviation).

TABLE 5.2: List of classifiers used in our study. For more details please refer the default parameter settings in WEKA [2].

Classifiers	**Summary of default parameters in WEKA**
SVM [302]	Kernel ='Polynomial', soft margin (C)=1.0
SL [313]	Number of boosting iterations=0
RF [303]	Initial number of random forests=100
MLP [315]	Learning rate=0.3, momentum=0.2, epochs=500
NB [309]	Without kernel estimator
BNet [312]	Hill climbing search algorithm
k-NN [329]	k=1, Euclidean distance
C4.5 [305]	Confidence threshold=0.25
ADTree [306]	Number of boosting iterations=10
M-V (b2) [317]	Combination rule=average probability
M-V (b3) [317]	Combination rule=average probability

5.4 Experimental Settings

This section presents the specification of the MR scanner and dataset used in this thesis.

5.4.1 Materials and Dataset

Our dataset consists of 418 T2-W MR images taken from 45 patients aged 54 to 74 (all patients had biopsy-proven prostate cancer). Each patient has between 6 to 13 slices covering the top to the bottom of the prostate gland. The prostate gland, malignant and transitional zone were delineated by an expert radiologist with more than 10 years experience in prostate MRI. All sequences with prostate cancer cases were confirmed malignancies based on TRUS biopsy reports. All malignant regions annotated cases were clinically significant cancer (Gleason score grade 7 and above). All patients underwent T2-W MR imaging at the Department of Radiology at the Norfolk and Norwich University Hospital, Norwich, UK. MR acquisitions were performed prior to radical prostatectomy. All images were obtained on a 1.5 Tesla magnet (Sigma, GE Medical Systems, Milwaukee, USA) using a phased array pelvic coil, with a 24×24 cm field of view, 512×512 matrix, $3mm$ slice thickness, and $0.5mm$ inter-slice gap. Approximately 60% of the source code was written in Matlab 2012a and 40% was written in Java. All experiments were run under the Windows 7 operating system with an Intel core i5 processor.

5.4.2 Training and Testing

All pixels within the radiologist's tumor annotation were extracted as prostate cancer samples (e.g. within the red outlined region in Figure 5.2). This area was truncated by the tumor mask, to ensure no pixels outside the tumor region were included in the malignant samples. On the other hand, every pixel outside the tumor region and within the PZ (under the magenta line in Figure 5.2) was considered as benign. Similarly, this region is truncated by the tumor and prostate gland masks to ensure no pixels within the tumor region and outside the prostate gland were included as benign samples. A stratified nine runs 9-fold cross-validation (9-FCV) scheme was employed. A leave one patient out approach was employed to ensure no samples from the same patient were used in the training and testing phases. We chose 9 folds instead of 10 folds to ensure each fold has the same number of patients (45 patients in our case, hence each fold contains 5 patients). Each classifier was trained and in the testing phase, each unseen instance/pixel from the testing data (taken from 5 randomly selected patients) was classified as malignant or non-malignant.

5.5 Experimental Results

In this study the experimental results are divided into four categories: performance based on classifiers, performance using different window sizes, performance evaluation before and after feature selection and feature evaluation (top 20 features).

5.5.1 Overall Performance

This section presents the overall performance of the proposed method using different classifiers. The performances were measured using the most popular metrics in the literature: Area Under the Curve (A_z, also known as AUC), Classification Accuracy (CA), sensitivity (Sen), specificity (Spe) and time taken for training and testing (t). A_z indicates the trade-off between the true positive rate against the false positive rate, where CA represents the number of pixel classified correctly. On the other hand, Sen measures the proportion of actual positives which are correctly identified (in this case the percentage of malignant pixels which are correctly identified) and Spe measures the proportion of actual negatives which are correctly identified. Sensitivity, specificity and accuracy can be calculated as $Sen = \frac{TP}{TP+FN}$, $Spe = \frac{TN}{TN+FP}$ and $CA = \frac{TP+TN}{TN+TP+FP+FN}$, respectively. TP and FP denote the number of true positives and false positives, respectively. Similarly, TN and FN indicate the numbers of true negatives and false negatives. The time taken by each classifier is measured in minutes to complete both training and testing in 9-FCV. On the other hand, p values indicate the significant difference for all metrics (A_z, CA and Sen) between M-V classifier (best 2) in comparison to the other 10 classifiers.

Table 5.3 shows that overall performances for each of the classifiers employed in this study and Table 5.4 contains the p values for the four main metrics between the M-V (best 2) classifier against the other 10 classifiers. The M-V (best 2) classifier outperformed all classifiers in all metrics and was statistically significant in terms of A_z with all individual classifiers (at least $p < 0.05$). The performance of M-V (best 2) classifier also significant for CA and Sen in comparison to most of the results produced by single classifiers. However, there is no significant difference in comparison to M-V (best 3) classifier for A_z, CA and Sen. These results are similar in terms of CA for BNet, ADTree and RF. The results indicate that using several classifiers in making decisions often produces better results due to their ability to handle complex data representation (or high dimensional data). For example, when the feature space

TABLE 5.3: Overall performances using different classifiers using a 11 × 11 sliding window.

Classifiers	$\boldsymbol{A_z}$(%)	$\boldsymbol{CA}$(%)	$\boldsymbol{Sen}$(%)	$\boldsymbol{Spe}$(%)	$\boldsymbol{t}$
M-V (b2)	**92.7 ± 7.4**	**85.5 ± 7.2**	**93.3 ± 9.1**	**86.5 ± 8.6**	21.70
M-V (b3)	92.3 ± 7.6	85.3 ± 7.6	92.7 ± 9.5	86.1 ± 8.1	24.30
BNet	90.0 ± 7.6	83.5 ± 7.6	90.8 ± 9.5	85.4 ± 7.9	3.83
ADTree	89.5 ± 8.9	83.7 ± 8.5	88.9 ± 11.6	85.6 ± 8.1	53.48
RF	87.6 ± 9.3	84.7 ± 7.2	91.8 ± 7.1	85.9 ± 8.3	19.22
MLP	87.4 ± 9.2	81.0 ± 9.9	86.5 ± 12.2	83.3 ± 9.3	294.17
NB	86.9 ± 10.5	80.5 ± 8.7	90.9 ± 10	82.6 ± 9.2	**1.23**
SL	85.6 ± 9.8	78.6 ± 9.5	80.1 ± 16.8	79.6 ± 10.2	444.37
C4.5	79.3 ± 9.8	82.4 ± 6.8	86.7 ± 7.6	82.3 ± 7.3	21.60
SVM	77.3 ± 9.9	78.5 ± 10.4	78.6 ± 17.2	80.5 ± 10.7	154.98
k-NN	68.8 ± 10.1	72.0 ± 11.9	81.9 ± 9.2	75.8 ± 9.4	116.78

TABLE 5.4: The p values for A_z, CA and Sen between the M-V classifier (best 2) against the other 10 classifiers from results in Table 5.3.

Classifiers	$\boldsymbol{A_z}$	$\boldsymbol{CA}$	$\boldsymbol{Sen}$	$\boldsymbol{Spe}$
M-V (b2)	-	-	-	-
M-V (b3)	0.8028	0.8965	0.7565	0.8181
BNet	0.0873	0.2005	0.2041	0.5287
ADTree	0.0643	0.2801	0.0455	0.6101
RF	0.0067	0.5961	0.3843	0.7339
MLP	0.0026	0.0135	0.0027	0.0910
NB	0.0024	0.0029	0.2340	0.0375
SL	0.0001	0.0001	<0.0001	0.0005
C4.5	<0.0001	0.0357	0.0002	0.0124
SVM	<0.0001	<0.0001	<0.0001	0.0034
k-NN	<0.0001	<0.0001	<0.0001	<0.0001

dimension is large many possible hypotheses could be created by a single classifier to build a prediction model (in our case the number of training samples can be up to 150,000 instances). This increases the probability that the classifier cannot guarantee finding the best hypothesis and approximation boundary of the target classes [97].

Hence, there is a risk of selecting a hypothesis or class boundary with a low accuracy on unseen data [97]. However, using several classifiers in making a decision increases the chance of selecting the best hypothesis by combining and averaging decisions or class boundaries for a final decision [97].

Individually, the BNet classifier performed best with A_z=90% followed by the ADTree, RF and MLP classifiers with an A_z=89.5%, 87.6% and 87.4%, respectively. BNet is expected to perform better than NB due to its ability of mapping the relationships among variables (or features) to build a predictive model without being restricted by the independence condition, whereas NB builds a predictive model based on a certain condition between two variables. Unfortunately, the independence condition is not always true and this leads to a less accurate model. MLP produced good results as it shares similar property with BNet (models the relationships among the input layer (which contains features), the hidden layer (the neuron) and the output layer (the actual class prediction)). On the other hand, RF and ADTree produced promising results due to their ability to perform like an ensemble classifier (consider various decisions and use averaging to improve predictive accuracy). In fact, ADTree employs boosting to improve the conventional decision tree algorithm (e.g. C4.5). Therefore ADTree produced better results in all three metrics compared to C4.5. In addition, RF and BNet contain efficient approaches for avoiding data over fitting. On the other hand, k-NN performed poorly in our study (A_z=67.5%). This may be caused by the number of neighbourhoods (k=1) as this restricts the algorithm to make decision based on a nearest single neighbourhood instead of based on several neighbours. Similarly, the low performance of SVM (A_z=76%) is expected to have been caused by default parameter settings. It is known that the performance of SVM is heavily affected by its parameters such as the choice of kernel, the kernel's parameters, and soft margin parameter C.

In terms of accuracy, there is still room for improvement (despite all predictive models produced CA>70%) as it can be seen that none of the predictive models managed to achieve CA>90%. The highest accuracy presented in Table 5.3 is achieved by the predictive models built by meta-classifiers via a voting approach which resulted in just above 85%. From a sensitivity point of view all predictive models achieved more than 80% except the model built by SVM. Both meta classifiers achieved more than 90% with acceptable time taken for training and testing below 25 minutes for 9-FCV. Individually, the predictive models built by BNet, RF and NB classifiers achieved more than 90% sensitivity in comparison to MLP (86.5%) and ADTree (88.9%), where

other predictive models achieved reasonable sensitivities. On the other hand, the NB classifier is the fastest predictive model taking less than 2 minutes to complete 9-FCV followed by the BNet classifier, which took less than 4 minutes. The slowest predictive models are the ones built by SL, MLP and SVM classifiers which took 444.37, 294.17 and 154.98 minutes, respectively.

Figure 5.6 shows the segmentation results produced by the proposed method in this chapter using different machine learning algorithms. In Figure 5.6, there are two malignant regions within the PZ which both are detected and segmented. The BNet classifier outperformed the other classifiers individually. However, the M-V (b2) combining the BNet and MLP classifiers produced the best A_z and CA. The M-V (b3) classifier (combination of the BNet, MLP and ADTree classifiers) also produced a good segmentation result but slightly poorer on the second malignant region. All classifiers successfully detected the first malignant region (left annotation in red) but a poor segmentation produced by the k-NN classifiers for both malignant regions. Note that the NB, ADTree and J48 classifiers produced good segmentation for the first malignant region and only the BNet classifier managed to segment most area (under the magenta line) of the second malignant region (right annotation in red). On the other hand, all classifiers produced false positives and false negatives for the second and first malignant regions, respectively.

Figure 5.7 shows another example of segmentation results with only one malignant region (red line) and has a visible appearance within the PZ. In comparison to the segmentation results in Figure 5.6, the NB classifier produced the best accuracy better than the M-V(b3) classifier. Nevertheless in this case, the M-V(b2) still produced the highest accuracy whereas the BNet, MLP, ADTree and SVM generated poor segmentation results.The M-V (b2) is a combination of the NB and J48 classifiers and an additional k-NN classifier for the M-V (b3).

Similar segmentation results were produced by M-V(b2) and M-V(b3) in another case in Figure 5.8. However, the RF, SL, MLP and SVM classifiers produced very poor results in comparison to the results produced in Figures 6.9 and 6.10. The BNet classifier remains the best classifier individually and the M-V (b2) remains the best classifier in comparison to the other machine learning algorithms employed in this study.

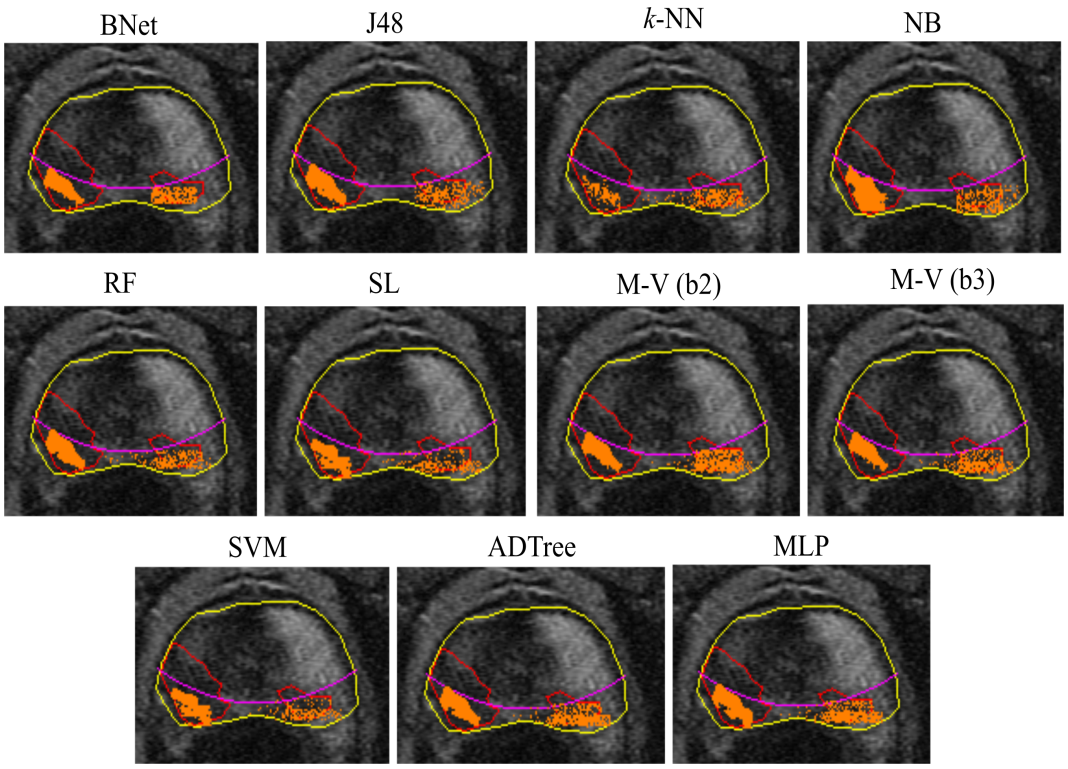

FIGURE 5.6: Example 1: Segmentation results using different machine learning algorithms.

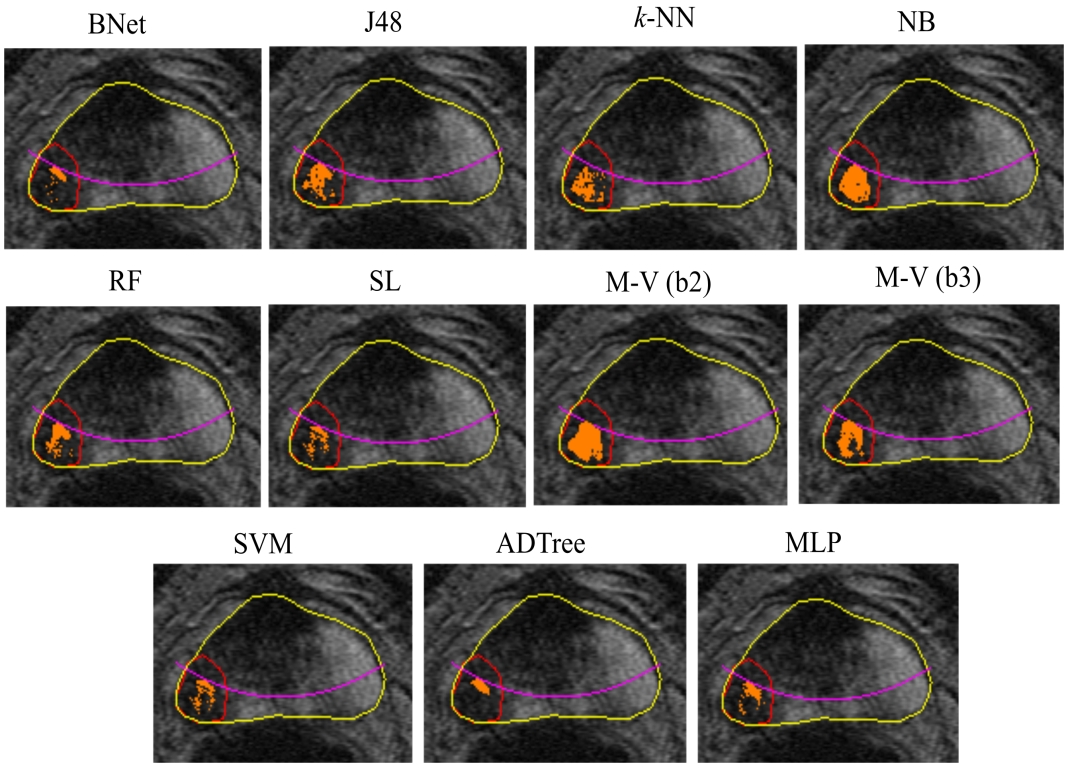

FIGURE 5.7: Example 2: Segmentation results using different machine learning algorithms.

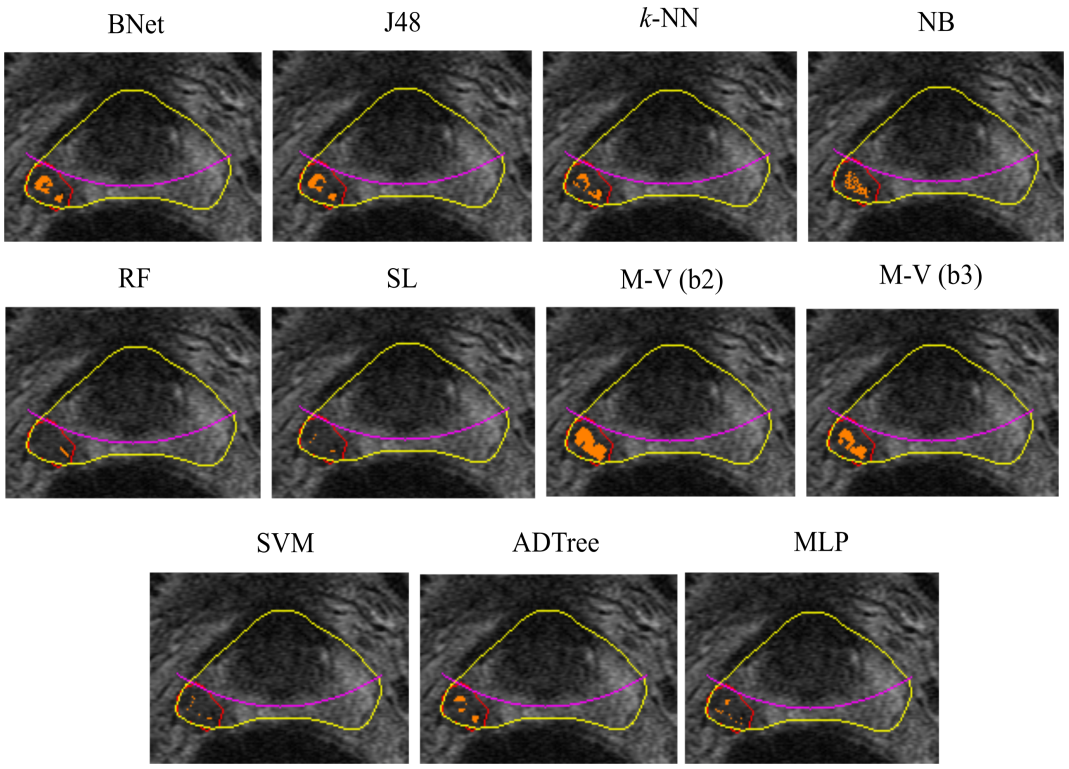

FIGURE 5.8: Example 3: Segmentation results using different machine learning algorithms.

5.5.2 Performance using Different Window Sizes

The selection of window size (ws) is one of the major issues in image processing. Many studies (such as [65, 231, 233, 263, 296]) in CAD-PC did not report how window sizes affect the overall performance of their methods. In this study, we investigated the effects of ws on performance quantitatively. For this purpose we conducted five experiments using the following ws: 3×3, 5×5, 7×7, 9×9, 11×11, 13×13, 15×15, 17×17 and 19×19. From a clinical point of view, these sizes enable us to see the variation of results according to the different sizes of malignant regions. We selected these sizes because these are the most common window sizes used in the literature. For this purpose we assess the performance of each classifier based on its A_z and CA values. Figure 5.9 and 5.10 are graphical representation for all classifiers employed in this study using different window sizes.

The results presented in Tables 5.5 and 5.6 show that ws affects the performance of the proposed method both in terms of A_z and CA values. In general using smaller ws such as 3×3 and 5×5 produced lower results compared to 9×9 and 11×11 due to insufficient information such as limited spatial information, limited intensities and grey level variations, the statistical values calculated from the neighbourhood are affected by

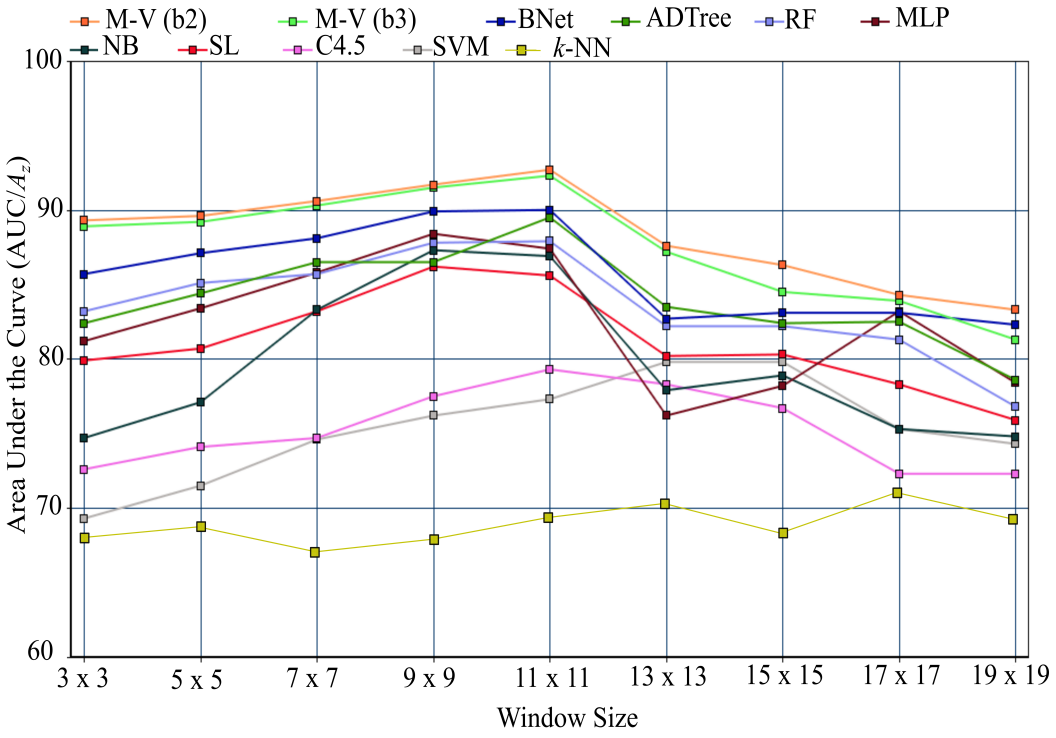

FIGURE 5.9: A graph representation for A_z values for all classifiers using different window sizes.

noise, etc. A study [330] suggests that large windows contain more information than small ones (hence, provides better texture characterisation). Moreover, several studies suggest that an appropriate guideline for window sizes is 7×7 and 9×9 [331, 332], where medium window sizes do not increase classification accuracy significantly [333] (in our case most CA and A_z values at 9×9 and 11×11 are very similar in Tables 5.5 and 5.6). Nevertheless, an absolute optimal size window is difficult to determine due to the complexity of the problem domain. As shown in Tables 5 and 6 the A_z values for all classifiers decreased after $ws = 11 \times 11$. A large ws causes the computed feature values within neighbouring pixels to over-represent the actual characteristic of the region.

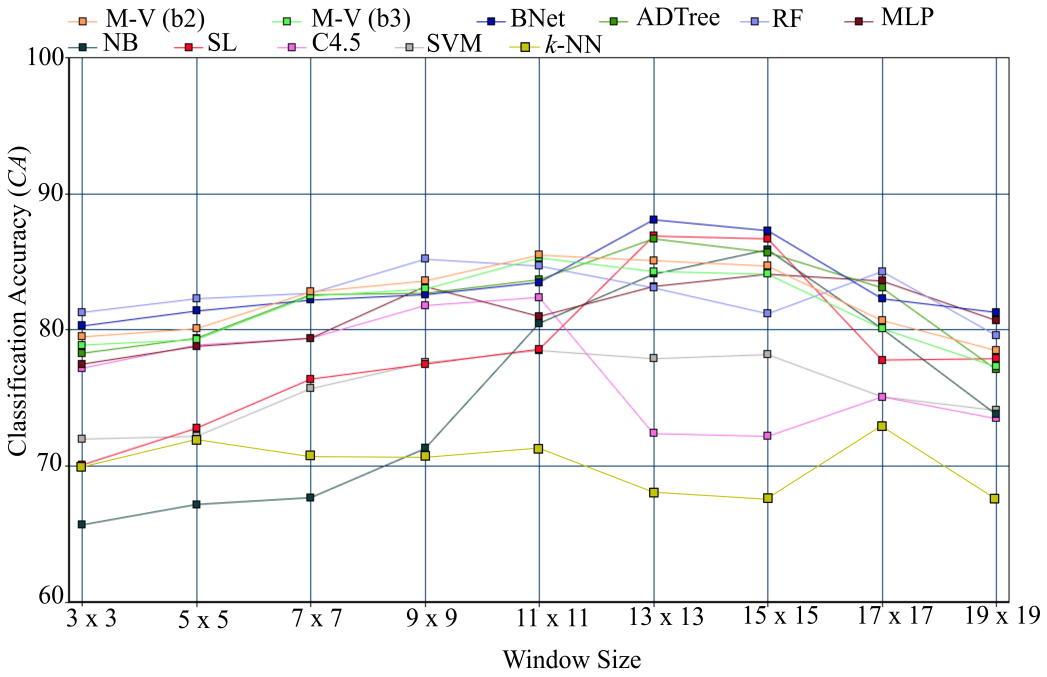

FIGURE 5.10: A graph representation for CA values for all classifiers using different window sizes.

TABLE 5.5: More details of A_z values using different window sizes

	3×3	5×5	7×7	9×9	11×11	13×13	15×15	17×17	19×19
M-V (b2)	89.3 ± 7.1	89.6 ± 7.3	90.6 ± 6.8	91.7 ± 6.8	**92.7 ± 7.4**	87.6 ± 6.9	86.3 ± 7.3	84.3 ± 6.9	83.3 ± 8.5
M-V (b3)	88.9 ± 7.5	89.2 ± 7.6	90.3 ± 7.3	91.5 ± 6.9	**92.3 ± 7.6**	87.2 ± 6.7	84.5 ± 7.1	83.9 ± 7.1	81.3 ± 9.1
BNet	85.7 ± 9.2	87.1 ± 9.8	88.1 ± 8.5	89.9 ± 9.9	**90.0** ± 7.6	**82.7** ± 9.9	83.1 ±11.6	83.1 ± 13.3	82.3 ± 14.9
ADTree	82.4 ± 9.5	84.4 ± 9.8	85.8 ± 10.3	86.5 ± 10.0	**89.5 ± 8.9**	83.5 ± 11.0	82.4 ± 11.6	82.5 ± 12.5	78.6 ± 17.0
RF	83.2 ± 8.1	85.1 ± 8.6	85.7 ± 9.0	87.8 ± 9.2	**87.9 ± 9.3**	82.2 ± 9.8	82.2 ± 11.9	81.3 ± 12.7	76.8 ± 15.4
MLP	81.2 ± 9.3	83.4 ± 9.7	85.8 ± 8.6	**88.4 ± 7.7**	87.4 ± 9.2	76.2 ±11.4	78.2 ± 13.1	83.2 ± 12.7	78.4 ± 14.3
NB	74.7 ± 10.3	77.1 ± 10.9	83.3 ± 10.5	**87.3 ± 9.6**	86.9 ± 10.5	77.9 ± 13.7	78.9 ± 13.4	75.3 ± 10.7	74.8 ± 11.4
SL	79.9 ± 8.3	80.7 ± 8.8	83.2 ± 8.9	**86.2 ± 7.1**	85.6 ± 9.8	80.2 ± 9.7	80.3 ± 11.9	78.3 ± 11.4	75.9 ± 11.9
C4.5	72.6 ± 9.2	74.1 ± 8.9	74.7 ± 9.5	77.5 ± 10	**79.3 ± 9.8**	78.3 ± 10.4	76.7 ± 12.5	72.3 ± 12.4	72.3 ± 14.4
SVM	69.3 ± 7.8	71.5 ± 7.6	74.6 ± 8.7	76.2 ± 7.9	77.3 ± 9.9	**79.8 ± 10.3**	79.8 ± 11.9	75.3 ± 11.7	74.3 ± 12.9
***k*-NN**	67.7 ± 7.7	68.5 ± 7.3	66.0 ± 8.2	67.5 ± 8.6	68.8 ± 10.0	**70.7 ± 15.2**	67.8 ± 17.7	71.9 ± 12.9	68.8 ± 11.0

This can be clearly seen in Tables 5.5 and 5.6 in both A_z and CA. A large ws (e.g. 17×17 and 19×19) caused both metrics to decrease below 85%. The results change between 5%-10% at $ws = 3 \times 3$ and 19×19. The proposed method achieved lower results at $ws = 3 \times 3$ due to insufficient information to characterise tumour regions. Similarly, at $ws = 19 \times 19$ the proposed method produced lowest results. According to Wolters *et al.* [98], the typical size ranges of malignant regions in prostate MRI are $5 - 20mm$ (on average $12.5mm$, using medium sizes ws are an appropriate (e.g. 9×9 or 11×11) in our case since it is the average tumor size and $ws = 3 \times 3$ would be too small). For most classifiers there are small improvement (less than 2%) for both A_z and CA using 9×9 and 11×11. Nevertheless, using different ws showed a significant difference for most classifiers in terms of A_z and CA.

TABLE 5.6: More details of CA values using different window sizes

	$\mathbf{3 \times 3}$	$\mathbf{5 \times 5}$	$\mathbf{7 \times 7}$	$\mathbf{9 \times 9}$	$\mathbf{11 \times 11}$	$\mathbf{13 \times 13}$	$\mathbf{15 \times 15}$	$\mathbf{17 \times 17}$	$\mathbf{19 \times 19}$
M-V (b2)	79.5 ± 7.3	80.1 ± 6.8	82.8 ± 7.1	83.6 ± 7.5	**85.5 ± 7.2**	85.1 ± 8.7	84.7 ± 8.1	80.7 ± 8.4	78.5 ± 10.5
M-V (b3)	78.9 ± 7.5	79.3 ± 7.3	82.5 ± 6.7	83.0 ± 8.1	**85.3 ± 7.6**	84.3 ± 8.3	84.1 ± 8.5	80.1 ± 8.1	77.3 ± 11.3
BNet	80.3 ± 6.1	81.4 ± 6.4	82.2 ± 6.6	82.6 ± 8.3	83.5 ± 7.6	**88.1 ± 9.8**	87.3 ± 11.5	82.3 ± 11.7	81.3 ± 15.2
ADTree	78.3 ± 6.1	79.4 ± 6.2	82.6 ± 6.3	82.7 ± 6.0	83.7 ± 8.5	**86.7 ± 11.6**	85.7 ± 13.7	83.1 ± 14.6	77.1 ± 18.0
RF	81.3 ± 5.9	82.3 ± 5.6	82.7 ± 6.0	**85.2 ± 5.3**	84.7 ± 7.2	83.1 ± 12.9	81.2 ± 15.4	84.3 ± 15.4	79.6 ± 18.7
MLP	77.5 ± 6.5	78.8 ± 6.3	79.4 ± 7.1	83.2 ± 6.3	81.0 ± 9.9	83.2 ± 11.5	**84.1 ± 13.9**	83.6 ± 14.7	80.7 ± 16.8
NB	65.7 ± 19.3	67.2 ± 20.4	67.7 ± 20.8	71.3 ± 19	80.5 ± 8.7	84.1 ± 11.9	**85.9 ± 11.6**	80.1 ± 12.7	73.8 ± 19.1
SL	70.1 ± 8.3	72.8 ± 8.2	76.4 ± 8.2	77.5 ± 9.1	78.6 ± 9.5	**86.9 ± 10.4**	86.7 ± 13.5	77.8 ± 15.8	77.9 ± 19.0
C4.5	77.2 ± 5.6	78.9 ± 5.4	79.4 ± 6.2	81.8 ± 6.1	**82.4 ± 6.8**	72.4 ± 13.7	72.2 ± 15.5	75.1 ± 17.0	73.5 ± 23.6
SVM	72.0 ± 8.5	72.2 ± 8.8	75.7 ± 9.2	77.6 ± 9.1	**78.5 ± 10.4**	77.9 ± 11.6	78.2 ± 12.5	75.1 ± 14.0	74.1 ± 16.5
***k*-NN**	69.9 ± 5.3	72.5 ± 5.7	70.5 ± 8.5	70.3 ± 12.4	72.0 ± 11.9	67.1 ± 13.7	66.4 ± 13.9	**73.8 ± 14.6**	67.3 ± 15.5

For example, increasing $ws = 3 \times 3$ to 11×11 increases CA value from 79.5 ± 7.3 to 85.5 ± 7.2 ($p < 0.0001$). However, some classifiers produced lower results when using a larger ws. For example the SVM classifier produced A_z=77.5% using 5×5 but produced 3% and 1% lower at 7×7 and 11×11, respectively. Similar for k-NN in terms of accuracy we can see that its best performance is using 5×5 with CA=72.5%. Interestingly, some classifiers produced the best CA at $ws = 13 \times 13$ such as BNet and ADTree. Nevertheless, most classifiers produced poorer CA at $ws \geq 15 \times 15$. Therefore, based on these results using $ws = 9 \times 9$, 11×11 or 13×13 is an appropriate guideline when selecting window size in our study.

TABLE 5.7: The p values between the best A_z value against the others across all classifiers and window sizes. Bold values indicate statistically insignificant.

	3 × 3	**5 × 5**	**7 × 7**	**9 × 9**	**11 × 11**	**13 × 13**	**15 × 15**	**17 × 17**	**19 × 19**
M-V (b2)	0.0264	0.0455	**0.1615**	**0.5028**	-	0.0007	< 0.0001	< 0.0001	< 0.0001
M-V (b3)	0.0155	0.0271	**0.1211**	**0.4237**	**0.8028**	0.0002	< 0.0001	< 0.0001	< 0.0001
BNet	< 0.0001	0.0022	0.0061	**0.0872**	**0.0873**	< 0.0001	< 0.0001	< 0.0001	< 0.0001
ADTree	< 0.0001	< 0.0001	0.0002	0.0008	**0.0643**	< 0.0001	< 0.0001	< 0.0001	< 0.0001
RF	< 0.0001	< 0.0001	< 0.0001	0.0054	0.0067	< 0.0001	< 0.0001	< 0.0001	< 0.0001
MLP	< 0.0001	< 0.0001	< 0.0001	0.0069	0.0026	< 0.0001	< 0.0001	< 0.0001	< 0.0001
NB	< 0.0001	< 0.0001	< 0.0001	0.0027	0.0024	< 0.0001	< 0.0001	< 0.0001	< 0.0001
SL	< 0.0001	< 0.0001	< 0.0001	< 0.0001	0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
C4.5	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
SVM	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
***k*-NN**	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001

Table 5.7 show the significance values between the best A_z against the others results for all classifiers and window sizes tested in the proposed method. The results indicate that combining classifiers produced significantly better A_z value in comparison to a single classifier. Nevertheless, there is no statistical significance when comparing it with the A_z values produced by the BNet classifier both at $ws = 9 \times 9$ and 11×11. Similarly, there is no statistical significance between the best A_z produced by M-V (b2) classifier against M-V (b3) classifier at $ws = 7 \times 7$, 9×9 and 11×11.

TABLE 5.8: The p values between the best CA value against the others across all classifiers and window sizes. Bold values indicate statistically insignificant.

	3 × 3	**5 × 5**	**7 × 7**	**9 × 9**	**11 × 11**	**13 × 13**	**15 × 15**	**17 × 17**	**19 × 19**
M-V (b2)	< 0.0001	< 0.0001	< 0.0032	0.0142	**0.1527**	**0.1235**	**0.073**	0.0001	< 0.0001
M-V (b3)	< 0.0001	< 0.0001	0.0016	0.0071	**0.1310**	0.0477	0.0384	< 0.0001	< 0.0001
BNet	< 0.0001	0.0001	< 0.0008	0.0041	0.0128	-	**0.7188**	0.0108	0.0117
ADTree	< 0.0001	< 0.0001	0.0015	0.0016	0.0226	**0.5352**	**0.3371**	**0.0561**	< 0.0001
RF	< 0.0001	0.0005	0.0016	**0.0801**	**0.0601**	0.0384	0.011	0.1615	0.0108
MLP	< 0.0001	< 0.0001	< 0.0001	0.0048	0.0006	0.0292	**0.1141**	**0.0872**	< 0.0001
NB	< 0.0001	< 0.0001	< 0.0001	< 0.0001	0.0001	**0.0818**	**0.332**	< 0.0001	< 0.0001
SL	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	**0.5754**	**0.5755**	< 0.0001	< 0.0001
C4.5	< 0.0001	< 0.0001	< 0.0001	< 0.0001	0.0013	< 0.0001	< 0.0001	< 0.0001	< 0.0001
SVM	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
***k*-NN**	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001

Table 5.8 presents the significance values between the best CA against the others results for all classifiers and window sizes tested in the proposed method. The best CA=88.1% by BNet classifier at $ws = 13 \times 13$, but there is no statistical significance difference if compared with the results produced by M-V (b2), ADTree, SL and NB at the same scale. At a slightly larger scale ($ws = 13 \times 13$), there is no statistical significance between the best CA with the result produced by the MLP classifier. On the other hand, at smaller scales (9×9 and 11×11), the RF classifier produced similar results with the best CA.

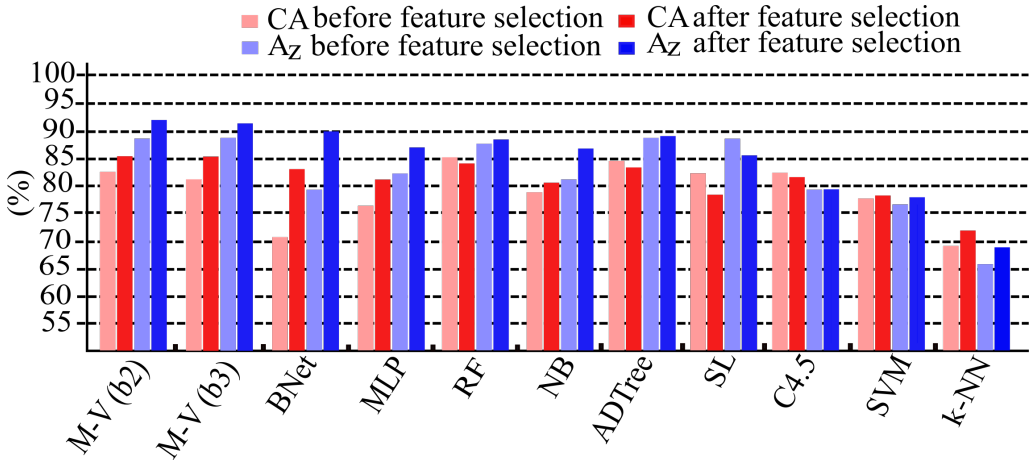

FIGURE 5.11: Performance comparisons among classifiers before and after feature selection using 11×11 window size

5.5.3 Performances before and after feature selection

A study by Niaf *et al.* [231] showed that feature selection can deliver a significant improvement in a classification model's performance. Therefore, in this study we investigated the effects (both for A_z and CA) of feature selection using two different window sizes (11×11). In the first experiment no feature was excluded and in the next experiment only features selected using CfsSubsetEval [327] were included in training and testing.

Figure 5.11 and 5.12 shows that there is a significant improvement in both A_z and CA ($p < 0.05$) for most classifiers after feature selection is performed due to different levels of data complexity. For example it is easier to build a predictive model in a lower dimensional dataset (e.g. 20 features per instance) than in a higher dimensional dataset (e.g. 215 features per instance). The difficulty in building a predictive model in a higher dimensional dataset increases the error rate (hence reduces predictive accuracy). However, the improvement for most classifiers is less than 3% except BNet, MLP, SL and NB classifiers which showed an improvement of approximately 5% to 10% in both A_z and CA. For example, the significant difference for A_z values before and after feature selection are $p < 0.0001$, $p = 0.0292$ and $p = 0.0072$ for BNet, MLP and NB classifiers, respectively at 11×11. This may be caused by two reasons:

1. BNet, MLP and NB classifiers share similar concepts of building a predictive model by mapping the relationships among features. Increasing the number of features means building a more complex predictive model (e.g. a large number

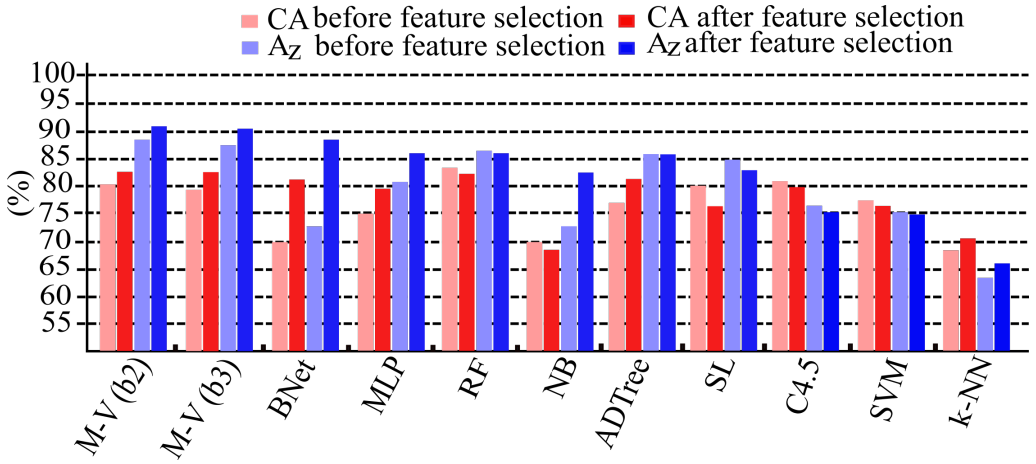

FIGURE 5.12: Performance comparisons among classifiers before and after feature selection using 7×7 window size

of features creates more complex network structure) as a larger number of relationships can be created. The complexity increases error rates and decreases the accuracy of the model in making a prediction in unseen cases.

2. BNet, MLP and NB classifiers do not have an approach to avoid data over fitting which means even weak and uncorrelated features will be considered in building a classification model.

Therefore performing feature selection is expected to be beneficial for BNet, MLP and NB classifiers as it decreases the level of complexity of the model. In comparison, for most tree-based classifiers only strong features will be considered in building the predictive model even without performing feature selection (the classifier selects the most correlated features in building a classification model). This suggests that the performance of the RF and ADTree classifiers are less affected by feature selection. From the results shown in Figure 5.11, differences for RF and ADTree classifiers before and after feature selection are less than 2% for both A_z and C_A. In addition, ADTree employs boosting procedures (uses many weak hypotheses to build a strong hypothesis) which often produce better results. Similarly, C4.5 is also a tree-based classifier which was not affected significantly by the feature selection (no significant difference $p = 0.4207$ for A_z at 11×11). On the other hand, the results produced by the Meta classifiers are mainly affected by the performance of the top three individual classifiers. In this study,the BNet, MLP, RF, NB and ADTree are among the top 3 classifiers combined in the M-V (based on the 3 best A_z values in the training phase).

5.5.4 Features Evaluation

We also investigated feature performance individually. For this purpose we conducted experiments by ranking the top 20 features based on the number of selection (ns) over the number of runs in 9-FCV (81 runs in total in this study) at nine different ws. The maximum value of ns is 81. The higher the ns the more frequently the feature has been selected by CfsSubsetEval [327].

TABLE 5.9: Top 20 most selected features using $ws = 3 \times 3$.

Features	$\boldsymbol{ns}$
Gaussian filter, Laplacian of Gaussian filter, image magnitude	81
Standard deviation, GLCM: sum of squares variance ($\theta = 135°$), GLCM: variance of homogeneities ($\theta = 0°, 45°, 90°, 135°$),	78
Local probability, local contrast, median	76
Variance, GLCM: Dissimilarity ($\theta = 45°$)	74
GLCM: variance ($\theta = 45°$), bar/spot filter ($(\sigma_x, \sigma_y) = (4, 12)$, $\theta = 150°$)	73
Bar/spot filter ($(\sigma_x, \sigma_y) = (2, 6)$, $\theta = 0°$)	72
bar/spot filter ($(\sigma_x, \sigma_y) = (4, 12)$, $\theta = 90°$),bar/spot filter ($(\sigma_x, \sigma_y) = (1, 3)$, $\theta = 0°$), edge filter ($(\sigma_x, \sigma_y) = (1, 3)$, $\theta = 90°$)	70
image magnitude of Sobel operator, mean	69
GLCM: Entropy ($\theta = 45°$), GLCM: Energy ($\theta = 90°$),	67

In our experiments evaluating the effects of ws on a feature's performance can be seen as one of the important parts in this study. Table 5.9 shows the list of top 20 most discriminant features based on the ns value for $ws = 3 \times 3$. There were 65 features selected with minimum ns=1 using $ws = 3 \times 3$. Features such as Gaussian filter, Laplacian of Gaussian filter and image magnitude were always selected by CfsSubsetEval [327] in 9-FCV of 81 runs. F_2 and F_6 features (see Table 1) dominate the list at the smallest window where none of Tamura's features were selected. F_1 features such as mean, median, local probability and local contrast are also among the most popular features at $ws = 3 \times 3$. On the other hand, 64 features were selected at least once at $ws = 5 \times 5$.

TABLE 5.10: Top 20 most selected features using $ws = 5 \times 5$.

Features	***ns***
GLCM: sum of squares variance ($\theta = 135°$), Gaussian filter, Laplacian of Gaussian filter, bar/spot filter ($(\sigma_x, \sigma_y) = (1, 3)$, $\theta = 30°$), bar/spot filter ($(\sigma_x, \sigma_y) = (1, 3)$, $\theta = 0°$), edge filter ($(\sigma_x, \sigma_y) = (1, 3)$, $\theta = 90°$), image magnitude, image magnitude of Sobel operator, variance	81
Image gradient $\theta = 90°$, image gradient $\theta = 0°$	80
Local probability, local contrast	79
Bar/spot filter ($(\sigma_x, \sigma_y) = (1, 3)$, $\theta = 150°$), bar/spot filter ($(\sigma_x, \sigma_y) = (4, 12)$, $\theta = 90°$)	77
Bar/spot filter ($(\sigma_x, \sigma_y) = (2, 6)$, $\theta = 90°$), bar/spot filter ($(\sigma_x, \sigma_y) = (2, 6)$, $\theta = 60°$), bar/spot filter ($(\sigma_x, \sigma_y) = (4, 12)$, $\theta = 150°$)	76
Bar/spot filter ($(\sigma_x, \sigma_y) = (2, 6)$, $\theta = 0°$)	75
Bar/spot filter ($(\sigma_x, \sigma_y) = (2, 6)$, $\theta = 30°$)	74

Results in Table 5.10 show that all features selected at least 80 times in 9-FCV come from F_1, F_2, F_5 and F_6, followed by local probability and local contrast with ns =79. Other features in the top 20 are a bank of bar/spot filters (also from F_6) covering different scales and orientations. Most of the features selected at $ws = 5$ are filter bank of Varma and Zisserman [1]. Increasing ws to 7×7 decreased the number of features selected to 59.

The results in Table 5.11 show that features from category F_5 and F_6 remain in favour. Variance along with image magnitude of Sobel operator remain among the most discriminant features. Interestingly, the ns value for GLCM: sum of squares variance ($\theta = 135°$) dropped to 68, placing it 16^{th} in the ranking. However, it can been seen that new features appear to be listed in the top 20 such as variance of cluster prominences ($\theta = 0°, 45°, 90°, 135°$), Tamura contrast and kurtosis. Moving up to $ws = 9 \times 9$, Tamura's contrast reached its best performance (ns=81) as it can be seen in Table 5.12. Similarly for image gradient of the Sobel operator, Gaussian filter and Laplacian of Gaussian filter remain among the top features with maximum ns value. The number of bar/spot filters has decreased and only one GLCM feature was selected in the top 20. New features such as percentiles 75% and edge filters appeared into the list while variance dropped its ns value from 81 to 75. Table 5.13 shows the results using $ws = 11 \times 11$. The results reveal a similar pattern for Gaussian filter, Laplacian of Gaussian filter, image magnitude of Sobel operator, Tamura contrast,

TABLE 5.11: Top 20 most selected features using $ws = 7 \times 7$.

Features	***ns***
Gaussian filter, Laplacian of Gaussian filter, bar/spot filter ($(\sigma_x, \sigma_y) = (1, 3)$, $\theta = 0°$), bar/spot filter ($(\sigma_x, \sigma_y) = (1, 3)$, $\theta = 90°$), image magnitude of Sobel operator, image magnitude , variance	81
Local contrast	80
Bar/spot filter ($(\sigma_x, \sigma_y) = (1, 3)$, $\theta = 30°$)	77
Bar/spot filter ($(\sigma_x, \sigma_y) = (1, 3)$, $\theta = 150°$)	74
Bar/spot filter ($(\sigma_x, \sigma_y) = (2, 6)$, $\theta = 0°$), local probability	73
GLCM: sum of squares variance ($\theta = 135°$)	68
Variance of cluster prominences ($\theta = 0°, 45°, 90°, 135°$)	67
Bar/spot filter ($(\sigma_x, \sigma_y) = (1, 3)$, $\theta = 90°$)	60
Tamura contrast	59
Image gradient ($\theta = 0°$)	55
Bar/spot filter ($(\sigma_x, \sigma_y) = (2, 6)$, $\theta = 60°$)	54
Kurtosis	52

image magnitude and variance. Although some ns values for some of the features (such as local probability, image gradient of Sobel operator and local contrast) dropped by at least 10, they remain in favour among the top 20 most discriminant features out of 216 features extracted in this study. Edge filters which were selected at $ws = 9 \times 9$ are not listed at $ws = 11 \times 11$, instead 3 GLCM features based on the feature's variance of four orientations ($\theta = 0°, 45°, 90°, 135°$) were selected.

Results in Tables 5.14 and 5.15 show that ns values for Gaussian and Laplacian of Gaussian have dropped to 76 and 70, respectively. These features are becoming less discriminant in much larger ws as shown in Tables 5.16 and 5.17 (but still in the top 20). Other features such as Tamura's contrast, image gradient, image magnitude and some edge filters remain consistent within the top 10 (note that the ns score for Tamura's contrast is always ≥ 80 when $ws \geq 9 \times 9$). Several features which were based on the variance of four orientation of single GLCM features (e.g GLCM: dissimilarities and autocorrelations) appear in the list at $ws \geq 17 \times 17$.

TABLE 5.12: Top 20 most selected features using $ws = 9 \times 9$.

Features	***ns***
Gaussian filter, Laplacian of Gaussian filter, bar/spot filter ($(\sigma_x, \sigma_y) = (4, 12)$, $\theta = 0°$), image magnitude of Sobel operator, image gradient of Sobel operator ($\theta = 45°$), image magnitude , image gradient ($\theta = 90°$), Tamura contrast	81
Image gradient of Sobel operator ($\theta = 90°$), local probability	80
Local contrast	79
Bar/spot filter ($(\sigma_x, \sigma_y) = (1, 3)$, $\theta = 0°$)	78
Image gradient of Sobel operator ($\theta = 0°$)	76
Variance	75
Percentiles 75%	74
Bar/spot filter ($(\sigma_x, \sigma_y) = (4, 12)$, $\theta = 45°$)	59
Image gradient ($\theta = 0°$)	42
An edge filter ($(\sigma_x, \sigma_y) = (2, 6)$, $\theta = 30°$)	36
An edge filter ($(\sigma_x, \sigma_y) = (2, 6)$, $\theta = 0°$)	29
GLCM: standard deviation of sum of variances ($\theta = 0°, 45°, 90°, 135°$)	24

Table 5.18 presents all the common features across different ws based on the total ns. The image magnitude has the highest $ns = 723$ (maximum $ns = 729$) followed by the image magnitude of Sobel operator ($ns = 687$). This indicate that these features are fairly consistent regardless to the ws (similar to the study claimed by Kovalev [70]). From these experiment results, it can be seen that the ws parameter affects the performance of the features. For example GLCM: sum of squares variance ($\theta = 135°$) appear to be in the top 20 list using smaller ws but did not perform well using larger ws (e.g 9×9 and 11×11). Similarly, Tamura's contrast was listed among the most discriminant feature using medium ws but performed poorly using smaller ws (5×5). In addition, some features performed well only using certain values of ws. For example Kurtosis appeared to be in the list only at $ws = 7 \times 7$, 13×13 and 15×15.

However, there are some features less affected by the ws parameter. For example the Gaussian filter, Laplacian of Gaussian filter, image magnitude of the Sobel operator, image magnitude, local contrast, local probability and variance (always in the top 20). Our experimental results support an earlier study conducted by Kovalev *et al.* [173] who claimed that image gradients have a more consistent behavior as a descriptor

TABLE 5.13: Top 20 most selected features using $ws = 11 \times 11$.

Features	***ns***
Gaussian filter, Laplacian of Gaussian filter, image magnitude of Sobel operator, Tamura contrast, image magnitude , variance	81
Bar/spot filter ($(\sigma_x, \sigma_y) = (2, 6)$, $\theta = 0°$), local probability	70
Bar/spot filter ($(\sigma_x, \sigma_y) = (1, 3)$, $\theta = 30°$)	63
Image gradient of Sobel operator ($\theta = 45°$)	54
Bar/spot filter ($(\sigma_x, \sigma_y) = (1, 3)$, $\theta = 150°$)	52
Bar/spot filter ($(\sigma_x, \sigma_y) = (1, 3)$, $\theta = 90°$)	51
Local contrast	50
GLCM: variance of autocorrelations ($\theta = 0°, 45°, 90°, 135°$)	48
GLCM: variance of sum of variances ($\theta = 0°, 45°, 90°, 135°$)	44
Percentiles 75%	43
GLCM: sum of variance ($\theta = 45°$)	41
An edge filter ($(\sigma_x, \sigma_y) = (1, 3)$, $\theta = 90°$)	36
GLCM: variance of cluster shades ($\theta = 0°, 45°, 90°, 135°$)	33
Image gradient ($\theta = 0°$)	30

TABLE 5.14: Top 20 most selected features using $ws = 13 \times 13$.

Features	***ns***
GLCM: sum of squares variance ($\theta = 135°$), bar/spot filter ($(\sigma_x, \sigma_y) = (1, 3)$, $\theta = 60°$), image gradient of Sobel operator ($\theta = 45°$), image magnitude, image gradient ($\theta = 90°$), Tamura contrast, variance and local contrast	81
Image magnitude of Sobel operator and local probability	78
Gaussian filter, Laplacian of Gaussian filter and image gradient ($\theta = 0°$)	76
GLCM: variance of cluster prominences ($\theta = 0°, 45°, 90°, 135°$) and percentile 75%	59
An edge filter ($(\sigma_x, \sigma_y) = (1, 3)$, $\theta = 0°$) and an edge filter ($(\sigma_x, \sigma_y) = (1, 3)$, $\theta = 90°$)	46
An edge filter ($(\sigma_x, \sigma_y) = (1, 3)$, $\theta = 150°$) and kurtosis	41
An edge filter ($(\sigma_x, \sigma_y) = (1, 3)$, $\theta = 30°$) and an edge filter ($(\sigma_x, \sigma_y) = (2, 6)$, $\theta = 90°$)	37

TABLE 5.15: Top 20 most selected features using $ws = 15 \times 15$.

Features	***ns***
Image gradient of Sobel operator ($\theta = 45°$), image magnitude, image gradient ($\theta = 90°$), Tamura contrast, variance	80
Image magnitude of Sobel operator, GLCM: variance sum of variance ($\theta = 0°, 45°, 90°, 135°$) and local probability	75
Gaussian filter, Laplacian of Gaussian filter, GLCM: variance of cluster prominences ($\theta = 0°, 45°, 90°, 135°$) and image gradient ($\theta = 0°$)	70
An edge filter ($(\sigma_x, \sigma_y) = (1, 3)$, $\theta = 0°$)	65
GLCM: sum of variance ($\theta = 0°$) and kurtosis	33
GLCM: sum of square variance ($\theta = 135°$)	29
An edge filter ($(\sigma_x, \sigma_y) = (2, 6)$, $\theta = 120°$)	26
GLCM: variance of sum of square variances ($\theta = 0°, 45°, 90°, 135°$), image gradient ($\theta = 90°$) and local contrast	24
Percentile 75%	14

TABLE 5.16: Top 20 most selected features using $ws = 17 \times 17$.

Features	***ns***
GLCM: variance of sum of variances ($\theta = 0°, 45°, 90°, 135°$), variance, Tamura's contrast, image gradient ($\theta = 0°$)	81
Image magnitude of Sobel operator and image magnitude	77
GLCM: variance of sum of square variances ($\theta = 0°, 45°, 90°, 135°$), local contrast	74
An edge filter ($(\sigma_x, \sigma_y) = (1, 3)$, $\theta = 60°$), image gradient of Sobel operator ($90°$) and local probability	73
Gaussian filter, Laplacian of Gaussian and An edge filter ($(\sigma_x, \sigma_y) = (2, 6)$, $\theta = 120°$)	43
GLCM: autocorrelation ($135°$)	33
GLCM: variance of cluster prominences ($\theta = 0°, 45°, 90°, 135°$)	23
GLCM: variance of dissimilarities ($\theta = 0°, 45°, 90°, 135°$), Variance of autocorrelations ($\theta = 0°, 45°, 90°, 135°$), image gradient ($0°$)	10

TABLE 5.17: Top 20 most selected features using $ws = 19 \times 19$.

Features	***ns***
Image gradient (0°), image magnitude and Tamura's contrast	80
GLCM: variance of sum of square variances ($\theta = 0°, 45°, 90°, 135°$), An edge filter ($(\sigma_x, \sigma_y) = (1, 3)$, $\theta = 60°$), variance and GLCM: variance of sum of variances ($\theta = 0°, 45°, 90°, 135°$)	70
Image gradient of Sobel operator ($\theta = 0°$) and image magnitude of Sobel operator	64
Local probability	58
Local contrast	49
Image gradient of Sobel operator	39
GLCM: variance of contrasts ($\theta = 0°, 45°, 90°, 135°$), Gaussian filter and Laplacian of Gaussian	22
GLCM: variance of dissimilarities ($\theta = 0°, 45°, 90°, 135°$)	17
GLCM: variance of cluster prominences ($\theta = 0°, 45°, 90°, 135°$)	10
GLCM: variance of autocorrelations ($\theta = 0°, 45°, 90°, 135°$) and An edge filter ($(\sigma_x, \sigma_y) = (2, 6)$, $\theta = 120°$)	9

compared to GLCM features. Tamura contrast, local probability, variance and percentiles 75% are among the most promising features but they need to be extracted with particular window sizes (e.g. 9×9 and 11×11). This is similar to the other features such as Gaussian and Laplacian of Gaussian (best at $ws \leq 13 \times 13$).

In a very large ws (e.g. 17×17 or 19×19), the ns value for Gaussian and Laplacian of Gaussian filters have decreased significantly from $ns = 76$ at $ws = 13 \times 13$ to $ns = 43$ ($ws = 17 \times 17$) and $ns = 22$ ($ws = 19 \times 19$), respectively. This may be caused using a larger size of filter will oversmooth (or blur) the textures within the prostate gland and remove important characteristics of benign and malignant tissues. Nevertheless, gradient based features such as image gradient and image magnitude showed consistent performance regardless to the size of ws used.

TABLE 5.18: Top 20 common features across different ws.

Features	$\boldsymbol{ns}$
Image magnitude	723
Image magnitude of Sobel operator	687
Local probability	668
Variance	623
Gaussian filter and Laplacian of Gaussian	616
Local contrast	601
Tamura contrast	543
Image gradient ($\theta = 0°$)	469
Image gradient ($\theta = 90°$)	403
Bar/spot filter ($(\sigma_x, \sigma_y) = (1, 3)$, $\theta = 0°$)	310
GLCM: sum of square variance $\theta = 135°$)	308
Bar/spot filter ($(\sigma_x, \sigma_y) = (2, 6)$, $\theta = 0°$)	290
Image gradient of Sobel operator ($\theta = 45°$)	242
Edge filter ($(\sigma_x, \sigma_y) = (1, 3)$, $\theta = 90°$)	233
GLCM: variance of sum of variances ($\theta = 0°, 45°, 90°, 135°$)	226
Bar/spot filter ($(\sigma_x, \sigma_y) = (1, 3)$, $\theta = 60°$)	224
Bar/spot filter ($(\sigma_x, \sigma_y) = (1, 3)$, $\theta = 30°$)	221
Image gradient of Sobel operator ($\theta = 90°$)	217
Bar/spot filter ($(\sigma_x, \sigma_y) = (1, 3)$, $\theta = 150°$)	203

Figures 5.13 and 5.14 show histogram representations for malignant and non-malignant data taken from four different features from the top 20 list at 9×9. On the other hand, Figures 5.15 and 5.16 show the features distribution in 2D feature space. Most malignant and benign data are quite separated in Gaussian versus Tamura contrast feature spaces. For image 2 in Figure 5.15, although there are fairly large amounts of data overlapping in local contrast versus percentiles 75%, the number of data separated are still quite noticeable. Similar in Figure 5.16 both feature distributions are visible despite there are large amounts of data overlapping.

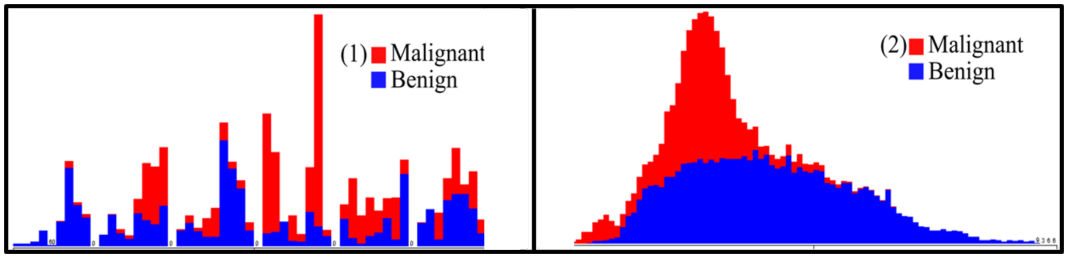

FIGURE 5.13: Histogram representations for some of the texture features used in this study. From left to right: GLCM : (1) local probability ($ns = 80$) , (2) percentiles 75% ($ns = 74$). Each graph represents 152, 445 instances taken at $ws = 9 \times 9$. Note that each histogram has the range of 0 to 1.

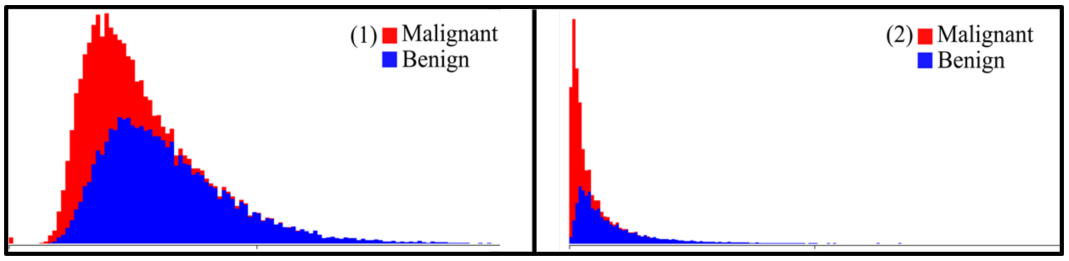

FIGURE 5.14: Histogram representations for some of the texture features used in this study. From left to right: GLCM : (1) local contrast ($ns = 79$) and (2) Tamura contrast ($ns = 81$). Each graph represents 152, 445 instances taken at $ws = 9 \times 9$. Note that each histogram has the range of 0 to 1.

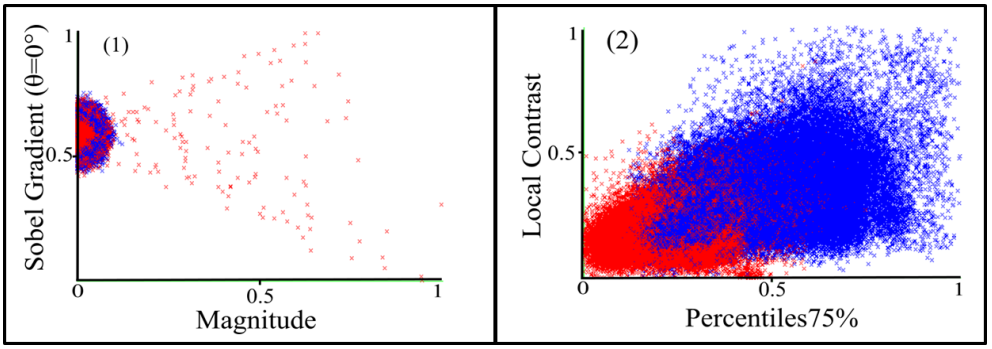

FIGURE 5.15: Pairwise scatterplots for two pairs of feature in 2D feature space. From left to right, (1) Gradient of the Sobel operator $\theta = 0°$) versus magnitude and (2) local contrast versus percentiles 75% (each graph represents 152, 445 instances taken using 9×9 sliding window).

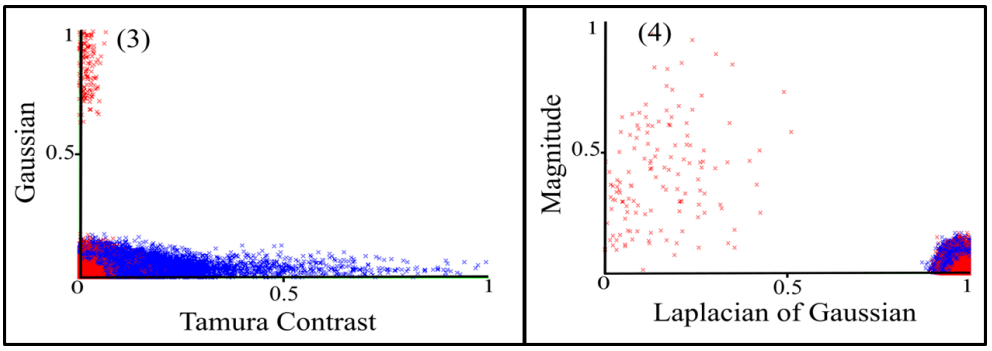

FIGURE 5.16: Pairwise scatterplots for two pairs of feature in 2D feature space. From left to right, (3) Gaussian versus Tamura contrast feature and (4) magnitude versus Laplacian of Gaussian (each graph represents 152, 445 instances taken using 9×9 sliding window).

5.6 Parameter Optimisation

Since the performance of most machine learning algorithms (classifiers) are depend on the parameters chosen by the users, we investigated the performance of three of the 11 classifiers employed in this study which are the k-NN, RF and the ADTree classifier. Performing parameter optimisation for each of the classifiers are time consuming due to the size of dataset (number of instances) and the complexity of the classifier itself (e.g. the SVM classifier has three main parameters and to find the best combination is time consuming). Note that in this section we used the data with features extracted using $ws = 11 \times 11$. Therefore the results presented in this section are preliminary results as we have not tested features extracted using different ws and classifiers. For the k-NN classifier we tested $k = 1$ up to $k = 41$ (with 2 neighbours interval). On the other hand, for the RF and ADTree classifiers, we tested the initial number of random forests (rF) from 5 to 165 (with an interval $rF = 5$) and the number of boosting (nB) from 1 to 41 (with an interval $nB = 2$, $nB = 10$ is included as the WEKA default value), respectively.

Figure 5.17 shows the A_z and CA results using different number of neighbours for the k-NN classifier. In terms of A_z, there is a visible difference between $k = 1$ and $k \geq 3$. At the larger number of neighbours ($k \geq 29$), the A_z slightly decreases to below 75%. On the other hand, there is a steady increment of CA (except at $k = 7$) as the number of neighbours increases before the same result achieved at $k = 39$ and $k = 41$. Figure 5.18 presents the performance of the RF classifier at different number of initial random forests. At $r \leq 25$, the RF classifier produced A_z<86% and CA<83%. The best A_z and CA are achieved at $rF = 90$ (also $rF = 115$ to $rF = 135$) and $rF = 100, 105$,

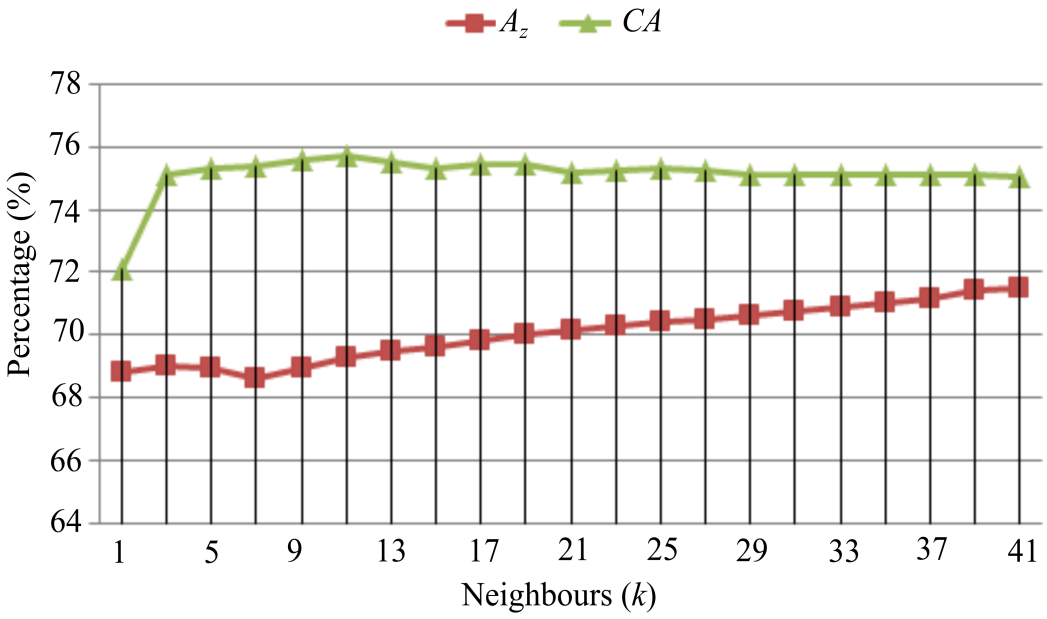

FIGURE 5.17: The A_z and CA values using different k values for the k-NN classifier. Default $k = 1$ in WEKA.

respectively. Similar A_z values were achieved at $rF \geq 140$, whereas CA decreased to approximately 83% at $rF \geq 160$.

Finally, Figure 5.19 shows the performance of the ADTree classifier using different numbers of boosting. The ADTree classifier produced the best A_z and CA at $nB = 10$ which is the default setting in WEKA. Other values of nB produced very similar results in both metrics expect $nB = 1$. Note that, a significant change in classification accuracy at $nB = 10$. This may be caused the training data has a better representative to the testing sample at $nB = 10$ as a larger nB (e.g. $nB > 10$) could possibly cause the predictive model is overfitting. These preliminary results show that the default parameters set in WEKA for ADTree and RF classifiers are good enough to get an idea how well a classifier works in a CAD-PC system. For example the ADTree classifier produced the highest A_z and CA at the default value given in WEKA. Similarly for the RF classifier the highest CA was achieved at $rF = 100$ which is the default value given in WEKA. Although the highest A_z was achieved at $rF = 90$ (and $rF = 115$ to 135), the difference is not significant at $rF = 100$ (around 1% difference). Nevertheless, around 3% difference between the default $k = 1$ given in WEKA and larger k (e.g. $k = 11$).

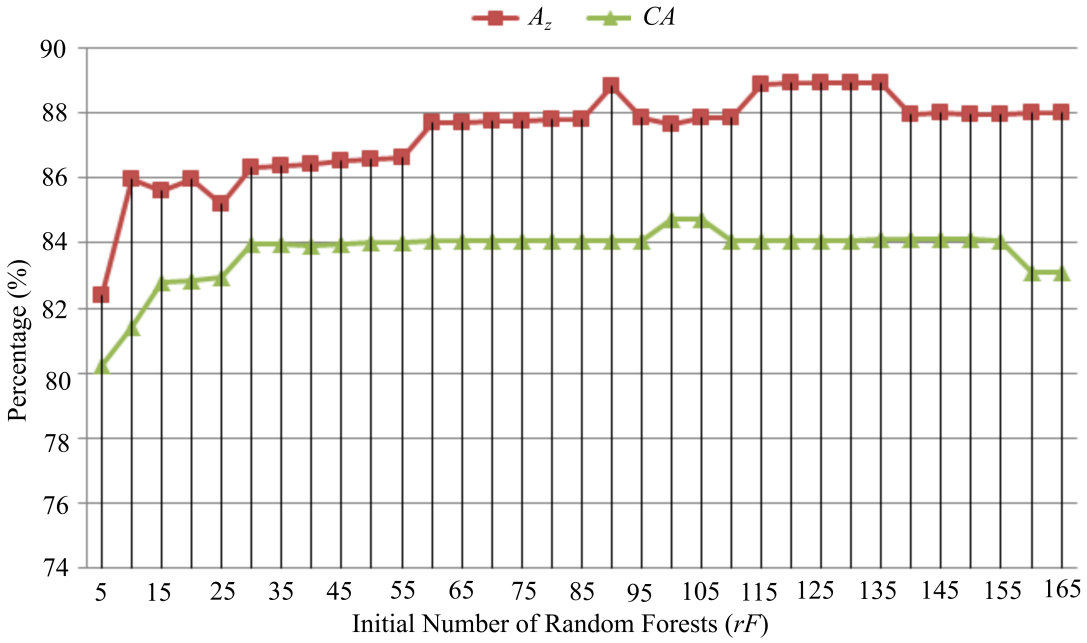

FIGURE 5.18: The A_z and CA values using different rF values for the RF classifier. Default $rF = 100$ in WEKA.

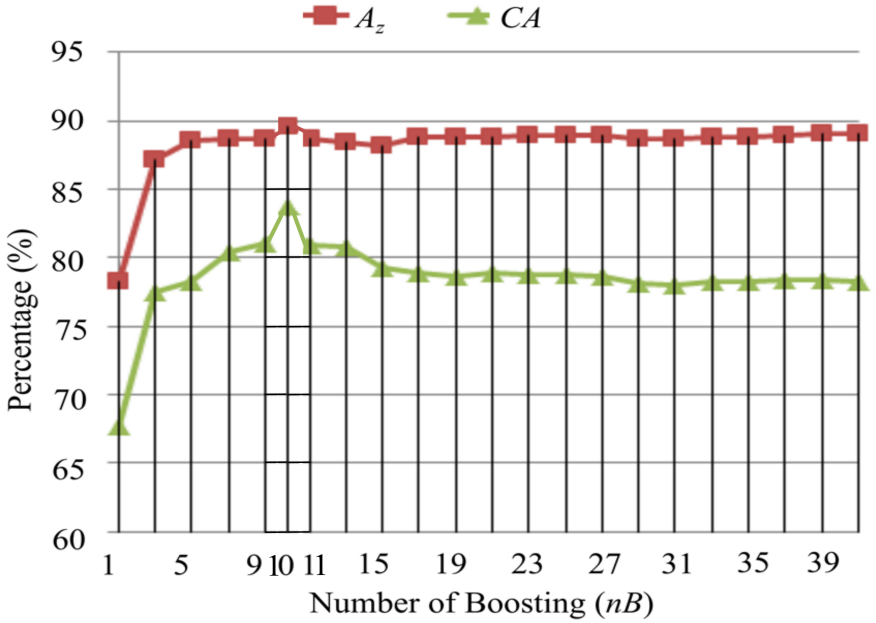

FIGURE 5.19: The A_z and CA values using different nB values for the ADTree classifier. Default $nB = 10$ in WEKA.

5.7 Discussion

The limitations of our study are: firstly, we are unable to compare our results quantitatively with the existing methods in the literature mainly due to the absence of public datasets. Every group of researchers has their own datasets which are not publicly available. This is currently one of the major issues in CAD-PC causing most studies in the literature to make only qualitative comparisons. Secondly, we are unable to compare our results quantitatively with the actual prospective clinical performance due to absence of radiologist performance in our dataset. Nevertheless, the studies in [46, 231] suggest that a radiologist performance ranging between A_z=80% to 86%. Although these values are based on their datasets, this roughly indicates that CAD-PC have the potential to assist radiologists as a second or first reader setting [46]. Thirdly, since our study employed 11 different classifiers to get the best possible results, performing parameter optimisation for each of the classifiers is time consuming and computationally expensive (therefore all parameters in this study were left on the default settings in WEKA). However, we have made an initial investigation regarding to this issue and included our preliminary results in Section 5.6. Finally feature's performance results are based on only one feature selection method (CfsSubsetEval [327]). Therefore the results maybe different using different feature selection methods which could be investigated in future work. However, from our experimental results we achieved several similarities with some studies [173, 231] in terms of feature consistency and performance particularly for F_5 features.

5.8 Summary

In summary, the proposed method extracted 215 texture feature from each pixel. Features are divided into six categories namely first-order statistical features, second-order statistical features, Tamura's features, gradient-based features and filter bank. Anisotropic diffusion was applied on each median-filtered image to efficiently remove noise and artifacts without blurring the image, hence preserving the edges and textures. Before feature selection is performed, features were normalised to [0,1] absolute values play a role. Subsequently, we employed the CfsSubsetEval [41] attribute evaluator and the GreedyStepwise search method in WEKA [2] to select only relevant features. Eleven classifiers were employed to build predictive models for the classification of the testing data. The evaluation results show that the proposed methods

achieved similar results to the state-of-the-art in the literature. This chapter highlights the following points:

1. CAD systems based on single modality T2-W MRI could achieve similar results to those based on multiparametric MRI. From a clinical point of view a CAD system based on T2-W MRI tends to have a lower running cost and is more convenient for both patient and physician.

2. Combining different classifiers produce better results in A_z especially when dealing with high dimensional data.

3. In this study feature selection improved the performance of the developed CAD system. This further supports an earlier study conducted by Niaf *et al.* [231].

4. We identified the most discriminant texture descriptors out of 215 features extracted in this study.

5. We investigated different ws and obtained optimal $ws = 11 \times 11$ (with very close results achieved at $ws = 9 \times 9$ and 13×13).

6. We performed an initial investigation of the effect of the parameters on the classifiers' performances as shown in Figures 5.19, 5.18 and 5.17.

In the next chapter, a texton based CAD approach will be proposed to classify benign and malignant tissues using the same machine learning techniques used in this chapter.

Chapter 6

A Texton Based Approach in Prostate Cancer Detection

This chapter proposes a CAD-PC methodology using textons to classify benign and malignant tissues within the prostate peripheral zone. For this purpose square patches are randomly extracted from malignant and benign regions. Subsequently, extracted patches are aggregated and clustered using k-means clustering technique to generate textons which represent both regions. All textons are combined to form the texton dictionary which was used to construct a texton map for every peripheral zone in the training images. Based on the texton map, histogram models for each malignant and benign tissue sample are constructed and used as a feature vector to train our classifiers. The proposed method was tested on 418 T2-W MR images taken from 45 patients and evaluation results show that the best three $A_z = 92.8\% \pm 5.9\%, 89.5\% \pm 7.1\%$ *and* $87.7\% \pm 8.6\%$ *achieved by Bayesian Network, Random Forest and C4.5 classifiers, respectively which are comparable with the state-of-the-art in the literature.*

6.1 Introduction

The ultimate goal of this chapter (and thesis) is to develop a CAD tool for prostate cancer detection within the PZ using a histogram based feature (e.g. textons). In 2013, Liu *et al.* showed that using histogram based features such as Local Binary Pattern (LBP) [74] and Histogram Orientated Gradient (HOG) [334] to discriminate benign and malignant regions in CAD-PC can achieve comparable results (A_z=83%)

with the other features in the literature. This was further supported by a recent study conducted by Kwak *et al.* [335] whose method used LBP and its variants in conjunction with b-value of DWI. In 2015, Lemaître *et al.* [62] conducted a review of CAD systems based on mono and multiparametric MRI for prostate cancer detection and reported that there are 42 studies in the literature from 2007 until 2014. Nevertheless, none of the methods have used textons to discriminate benign and malignant tissue in their studies. Although the term texton was first introduced in the 80's, it did not get much attention until a study of texture classification by Leung and Malik [159] in 2001. Later, similar studies showing promising results in texture classification were conducted by Varma and Zisserman [1, 168] in 2005 and 2009, respectively. On the other hand, in medical image analysis textons have been used in retinal vessel segmentation [171] and lung cancer detection [172].

As previously discussed in Chapter 2, due to the lack of a good mathematical model the term 'texton' remains unclear in the literature [170]. However, textons can be seen as a representative of micro-structures in natural images and are considered as the atoms of pre-attentive human visual perception [169]. Many texton based approaches [139, 159, 336–339] in image classification used histograms as representatives of the texture in an image. In the original approach, textons are represented by means of a collection of filter bank responses obtained from large filter banks such as the MR8 [139], LM [159], S [340] filter banks and Gabor filter [341]. All the response vectors are collected and clustered using k-means and the resulting cluster centers are called textons (hence, in a simplest definition textons are the k-means' cluster centers). Nevertheless, the study in [161] showed textons can be generated by directly clustering the image's pixel values from patches without the need of filter banks (hence, speeding up the process of constructing the texton dictionary) and more robust in characterising textures.

The study in [342] suggested there are three reasons for the relatively strong performance of textons generated from the image's pixels in comparison to textons generated from the convolved image's pixels. First, the use of filter banks reduces the number of textons that can be extracted from a texture image [336]. For example, the number of textons are significantly reduced when an image of 250×250 pixel is convolved with a 50×50 filter. This affects the quality of the histogram models to characterise a particular texture (e.g. insufficient information to model the actual representation of the texture). Second, the large number of filter banks leads to errors in the edge localisations which may significantly change the geometry of the textons, leading to

errors in the estimation of the texton frequency histogram [336]. Finally, most filter banks lead to some blurring on the texture which might remove local details in the texture hence alter the actual texton representation of the texture [336]. Motivated from these suggestions, the proposed method in this chapter did not use any filter banks but took the image's pixels directly to generate textons.

The novel contributions of this chapter are:

1. This is the first CAD method which has investigated the use of textons in classifying benign and malignant tissues within the prostate.
2. The proposed method learns directly from image pixels without the need of using a filter bank. In comparison, most CAD-PC in the literature compute large numbers of texture descriptors, which are computationally expensive. In fact, computing a large number of texture descriptors also leads to an additional essential step such as feature selection or dimensionality reduction.

6.2 An Overview of a Texton Based Approach

Figure 6.1 shows a typical texton based approach in texture image classification. To generate histogram models representation of a particular texture class, texture images from the same class are filtered (e.g. convolution) by a filter bank (e.g. Gabor filter, LM, MR8, etc.) resulting in a set of filtered images. The resultant filter responses are aggregated and clustered using the k-means algorithm to generate textons which

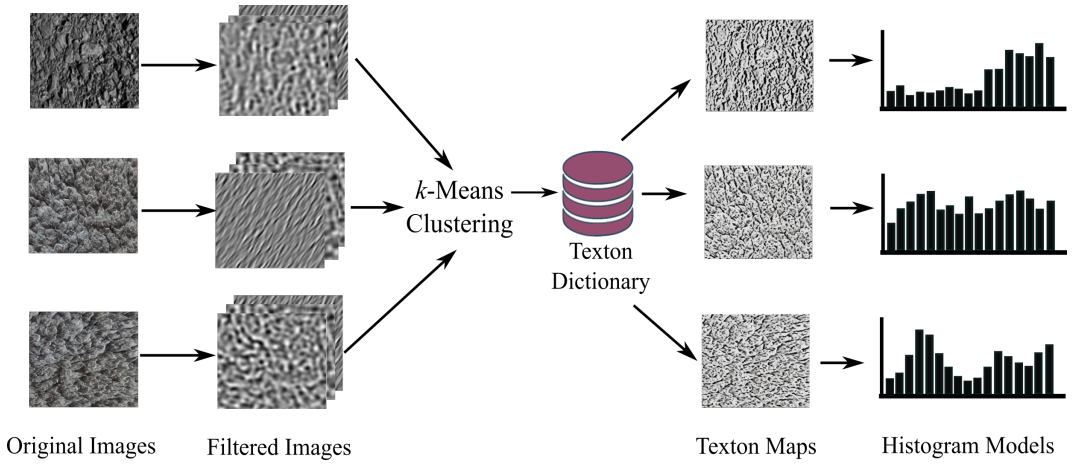

FIGURE 6.1: A general overview of a texton based approach in image classification adapted from [1]. The original images are the training set of the same texture class.

are stored in $m \times n$ matrix called 'texton dictionary'. Each texton is unique and has its own identity (id). To construct a texton map for each texture image, the image is first convolved with a filter bank and then labelling (using texton id) each filter response with the texton which is the most similar or closest (e.g. euclidean distance) with the texton in the 'texton dictionary'. In this process each pixel in the image is replaced with the corresponding texton id. The histogram model for an image represents the frequencies of texton ids in the texton map. The same workflow is applied for the other texture images. This means, if there are 100 texture images in the particular class, we will have 100 histogram models which represent the different texture variations of the class. In the testing phase, the same procedure is applied to generate the histogram model of the test image. Finally, the histogram model is compared (using similarity measure e.g. X^2) with each of the histogram models from all texture classes. As we can see, the speed of typical texton based approaches are affected by the number of convolutions in the bank filter (e.g. the number of filters).

6.3 Methodology

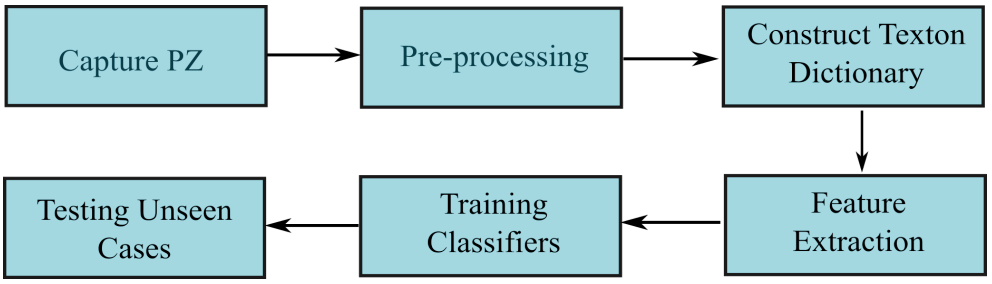

FIGURE 6.2: A general overview of the proposed method.

Figure 6.2 shows a flowchart of our proposed method. For every input image, we roughly estimate the area of the prostate's PZ followed by normalisation and noise reduction. Subsequently, for every training image we randomly extract patches from benign and malignant regions within the PZ and employed k-means clustering to generate textons (the texton dictionary is constructed). Each pixel in every training image is labelled with the texton to which which it lies closest, producing texton maps. Using the texton maps, a histogram of textons (the frequency of each texton occurs) is constructed for every pixel within the PZ . All histograms of textons from all pixels are treated as feature vectors and used to train our classifiers. Finally, at the testing phase, every unseen PZ is processed in the same way and the trained classifiers are used to decide, for each pixel, whether it belongs to the benign or malignant class.

6.3.1 Capturing the Peripheral Zone

To capture (or segment) the peripheral zone we employed exactly the method previously used in Chapters 4 and 5. Figure 6.3 shows example MRI images with the

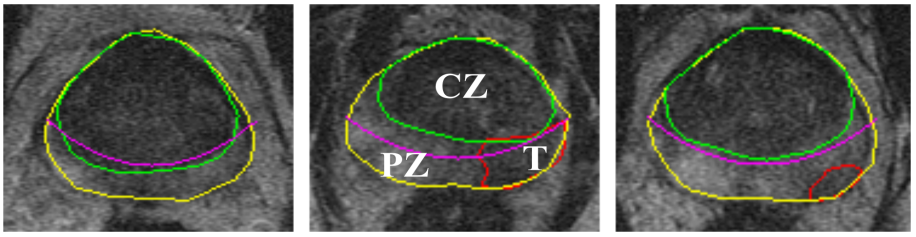

FIGURE 6.3: Example images of prostate MRI with ground truth delineated by an expert radiologist and the estimated PZ region under the magenta line.

ground truth of prostate gland, central zone (CZ) and tumor (T) represented in red, yellow and green, respectively, while magenta line is the estimated boundary of PZ.

6.3.2 Pre-processing

Most of the CAD systems in the literature integrate a pre-processing step as part of the methodology. It is an important step to remove unwanted artifacts or noise which can affect the classification accuracy due to alteration of the actual texture's representation. In this method we employed a robust noise reduction method of Perona and Malik [318] called 'anisotropic diffusion filtering' which was explained in Chapter 5. Note that the anisotropic diffusion filter was applied on median filtered images as described in [192, 204, 205] to preserve edge boundaries. Subsequently, image intensities were normalised to zero mean unit variance and anisotropic diffusion filtering [205, 318] was applied to remove noise.

6.3.3 Texton Dictionary

Figure 6.4 shows the workflow to construct the texton dictionary. Textons (e.g. 9×9 square window) were retrieved from benign and malignant regions. To construct the texton dictionary, we followed the studies of Varma and Zisserman [1, 139, 161, 168, 342]. For every PZ area in the training images we randomly extract 9×9 patches of raw pixels from benign and malignant regions. Subsequently, all patches extracted from benign regions were aggregated and clustered into textons using the k-means

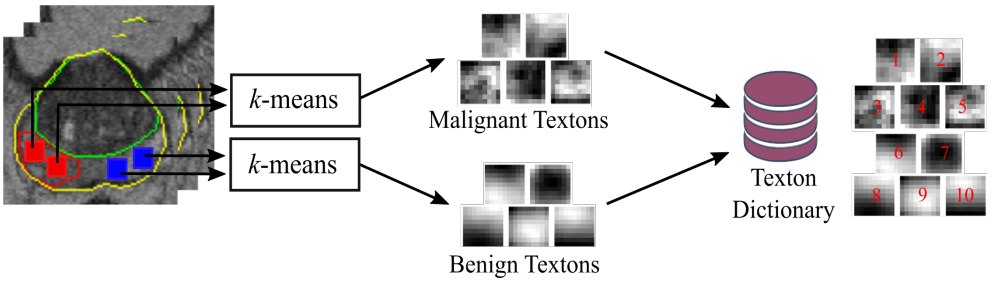

FIGURE 6.4: Generating the texton dictionary. Patches from the same class are aggregated and clustered using the k-means algorithm. Blue and green patches are malignant and benign samples, respectively.

algorithm. The same process was performed for all patches extracted from malignant regions. To summarise the k-means algorithm [343–345]:

1. k points are selected randomly in the feature space which represent the initial group centroids of k clusters.
2. Each object in the feature space is assigned to the cluster with the closest centroid.
3. Recalculate the position of the k centroids when all objects are assigned.
4. Repeat Steps 2 and 3 until the centroids no longer move (until it converges).

The cluster centroids produced by the k-means algorithm are the textons. Once all textons from both classes (benign and malignant) were generated, they were combined to form the texton dictionary. As shown in Figure 6.4, each texton is unique and has its own id ($TX = tx_1, tx_2, tx_3.....tx_n$) saved in a matrix which will be used to construct the texton map for each image.

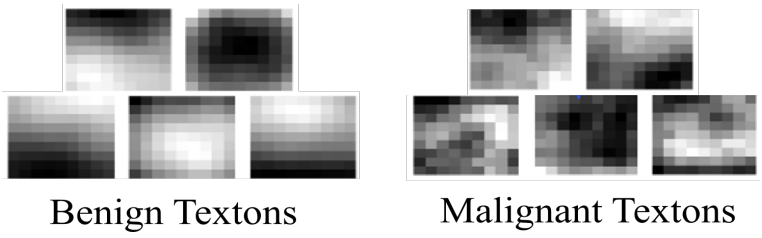

FIGURE 6.5: Example of textons generated from benign (bottom row) and malignant (top row) regions.

Figure 6.5 shows a sample of textons extracted from both classes. Visually it can be seen that textons generated from benign regions look smoother than the ones retrieved from malignant regions.

6.3.4 Feature Extraction

In this study, feature extraction is a process to generate histogram models for benign and malignant tissues. This phase consists of two main stages. The first stage (Figure 6.6) is to generate the texton map, where every pixel in the image (within the PZ) is assigned to the closest texton (using euclidean distance) in the texton dictionary. We used a sliding window W^{T} (blue patch in Figure 6.4) of the same size as the previous phase and finding the shortest Euclidean distance between W^{T} with each texton in TX.

Subsequently, the central pixel in W^{T} is replaced with the texton *id* which is the closest to W^{T}. This process is repeated for every pixel in the image until all pixels are assigned with the corresponding textons '*ids*'. By the end of this stage, a texton map is constructed for every PZ which will be used in the subsequent stages. Figure 6.7 shows examples of texton maps of three PZs generated in this phase. Each pixel within the PZ was replaced with the corresponding texton 'id'.

At the second stage, using the texton map we are able to generate a histogram model for each pixel by using a sliding window of the same size in the previous stage. A histogram for each pixel is constructed based on the occurrence of each texton's frequency within the neighborhood of the central pixel (including the central pixel). Figure 6.8 shows an example of constructing a histogram for each pixel. In this example, there are 10 textons (5 textons for each class) in the dictionary and each histogram of a tissue is constructed based on 5×5 window size; this means the histogram is constructed based on the texton frequency within 25 pixels (or 81 pixels for a 9×9 window size). Note that every histogram is normalised to unity. This yields a histogram for each tissue in the training image which is used as a feature vector representing every pixel. To this end, each pixel is represented in a *txt* dimensional feature space where *txt* is

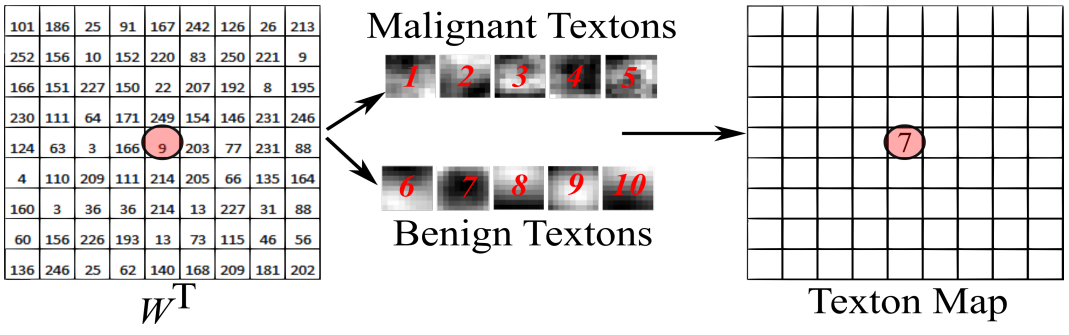

FIGURE 6.6: A graphical illustration on how to construct a texton map of an image.

FIGURE 6.7: Examples of texton maps of three PZs taken from three different prostates.

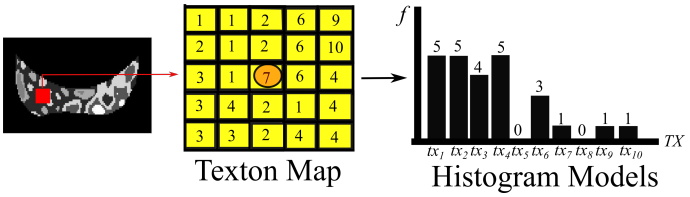

FIGURE 6.8: Constructing a histogram for each pixel from the texton map.

the number of textons in the dictionary (10 in this example). Similarly, if there are 30 textons (15 textons per class), each pixel is represented in a 30 dimensional feature space. It should be noted that the data dimension is independent of W^{T}.Finally, the constructed histogram for each pixel will be treated as feature vector in the training and testing phases.

6.4 Experimental Settings

This section presents the specification of the MR scanner and dataset used in this thesis.

6.4.1 Materials and Dataset

Our dataset consists of 418 T2-W MR images taken from 45 patients aged 54 to 74 (all patients had biopsy-proven prostate cancer). Each patient has between 6 to 13 slices covering the top to the bottom of the prostate gland. The prostate gland, malignant and transitional zone were delineated by an expert radiologist with more than 10 years experience in prostate MRI. All sequences with prostate cancer cases were confirmed malignancies based on TRUS biopsy reports. All malignant regions annotated cases were clinically significant cancer (Gleason score grade 7 and above). All patients underwent T2-W MR imaging at the Department of Radiology at the Norfolk and Norwich University Hospital, Norwich, UK. MR acquisitions were performed prior to

radical prostatectomy. All images were obtained on a 1.5 Tesla magnet (Sigma, GE Medical Systems, Milwaukee, USA) using a phased array pelvic coil, with a 24×24 cm field of view, 512×512 matrix, $3mm$ slice thickness, and $0.5mm$ inter-slice gap. Approximately 60% of the source code was written in Matlab 2012a and 40% was written in Java. All experiments were run under the Window 7 operating system with an Intel core i5 processor.

6.4.2 Training and Testing

All pixels within the radiologist's tumor annotation were extracted as prostate cancer samples. This area was truncated by the tumor mask, to ensure no pixels outside the tumor region were included in the malignant samples. On the other hand, all pixels outside the tumor region and within the PZ were considered benign samples. Similarly, this region is truncated by the tumor and prostate gland masks to ensure no pixels within the tumor region and outside the prostate gland were included as benign samples. A stratified nine runs 9-fold cross-validation (9-FCV) scheme was employed. The folds were populated on a patient basis to ensure no samples from the same patient were used in the training and testing phases. We chose 9 folds instead of 10 folds to ensure each fold has the same number of patients (45 patients in our case, hence each fold contains 5 patients). Each classifier was trained and in the testing phase, for every unseen instance/pixel from the testing data (taken from 5 randomly selected patients) was classified as malignant or non-malignant.

6.5 Experimental Results

In this study the experimental results are divided into three categories: performance based on classifiers, performance using different window (or patch) sizes and performance using different number of textons. The performances were measured using the most popular metrics in the literature: Area Under the Curve (A_z, also known as AUC), Classification Accuracy (CA), Sensitivity (Sen) and Specificity (Spe). A_z indicates the trade-off between the true positive rate against the false positive rate, where CA represents the number of pixel classified correctly. On the other hand, Sen and Spe measure the proportion of actual positives and negatives which are correctly identified, respectively (in this case the percentage of malignant pixels which are correctly identified). Sensitivity, specificity and accuracy can be calculated as

$Sen = \frac{TP}{TP+FN}$, $Spe = \frac{TN}{TN+FP}$ and $CA = \frac{TP+TN}{TN+TP+FP+FN}$, respectively. TP and FP denote the number of true positives and false positives, respectively. Similarly, TN and FN indicate the numbers of true negatives and false negatives. On the other hand, p values indicate the significant difference for all metrics (A_z, CA and Sen) between the best classifier in comparison to the other classifiers.

6.6 Overall Results

This section summarises the overall results. Table 6.1 and 6.2 show that best A_z, CA, Sen and Spe for all classifiers regardless of the ws and txt. Overall, BNet classifier with A_z=92.8% and CA=84% outperformed the other classifiers in both metrics, which is statistically significant against all classifiers except the CA of the RF classifier. This means, RF classifier produced the second best results in both metrics with A_z=89.5% and CA=81.1% at $ws = 11 \times 11$ and $txt = 6$. The lowest A_z was produced by the NB classifier with just above 82%. In comparison, all classifiers achieved A_z>82% but in terms of accuracy most classifiers achieved CA<80%. On the other hand, the overall results show that best results for all classifiers employed were achieved using either $ws = 9 \times 9$ or 11×11 and $txt = 6$ or 16. Considering the best A_z values of all classifiers, results suggest that the proposed method can achieve similar performances to the other CAD-PC in the literature. Furthermore, based on the best A_z, our method qualitatively outperformed most of the existing methods. Nevertheless, in terms of accuracy there is space for improvement. Table 6.3 presents the statistical significance in a hypothesis test between the best A_z and CA against the other results regardless of the ws and txt.

Figure 6.9 shows the segmentation results produced using different classifiers. The RF and MLP classifiers produced good accuracy covering most area of the malignant region. In comparison to the features used in Chapter 5, the k-NN and SVM classifiers performed better than the NB and ADtree classifiers in this case.

Figure 6.10 shows another example of segmentation results in a more difficult case (the malignant region is obscure and the size of the PZ is small). In this example, the RF classifier again produced the highest accuracy followed by the k-NN classifier. The BNet and NB classifiers generated reasonable segmentation results but poor segmentation results of the SVM and SL classifiers with both have large false positives.

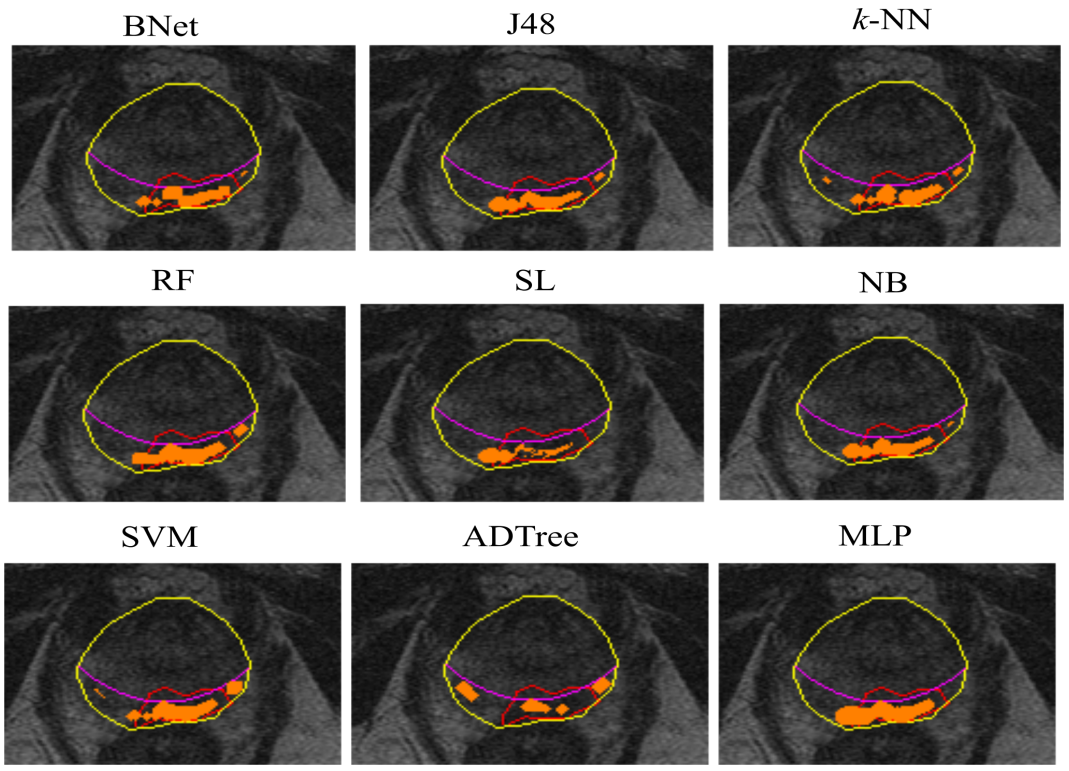

FIGURE 6.9: Example 1: Segmentation results using different machine learning algorithms.

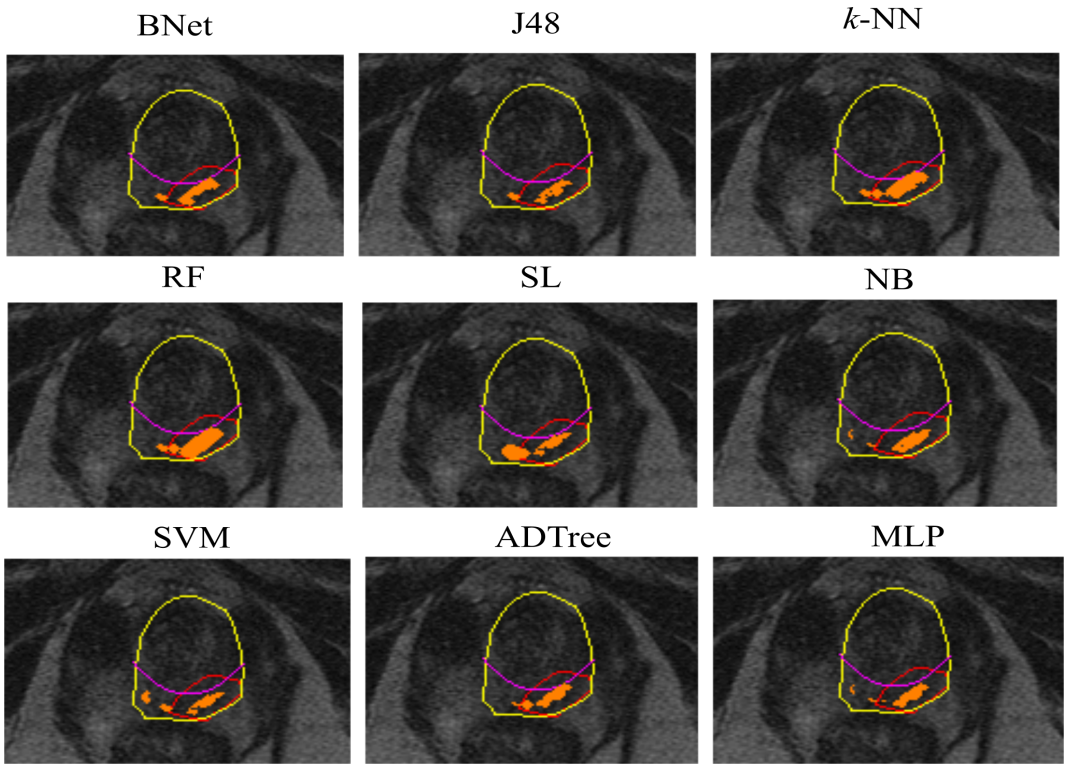

FIGURE 6.10: Example 2: Segmentation results using different machine learning algorithms.

TABLE 6.1: The best A_z and CA for all classifiers regardless of the ws and number of textons (txt) in the texton dictionary.

	$\boldsymbol{A_z}$ (%)	$\boldsymbol{CA}$ (%)
ADTree	83.8 ± 11.9 ($ws = 9 \times 9, txt = 6$)	75.3 ± 13.6 ($ws = 11 \times 11, txt = 16$)
BNet	**92.8 ± 5.9** ($ws = 11 \times 11, txt = 16$)	**84.0 ± 7.0** ($ws = 11 \times 11, txt = 16$)
C4.5	87.9 ± 8.6 ($ws = 11 \times 11, txt = 6$)	78.6 ± 10.5 ($ws = 11 \times 11, txt = 6$)
k-NN	86.9 ± 7.5 ($ws = 9 \times 9, txt = 6$)	80.2 ± 9 ($ws = 11 \times 11, txt = 6$)
MLP	85.6 ± 9.3 ($ws = 11 \times 11, txt = 16$)	76.2 ± 11.5 ($ws = 11 \times 11, txt = 16$)
NB	82.5 ± 11.1 ($ws = 11 \times 11, txt = 16$)	76.7 ± 15.4 ($ws = 11 \times 11, txt = 16$)
RF	89.5 ± 7.1 ($ws = 11 \times 11, txt = 6$)	81.1 ± 9.3 ($ws = 11 \times 11, txt = 6$)
SL	81.4 ± 10.6 ($ws = 9 \times 9, txt = 24$)	75.7 ± 14.2 ($ws = 11 \times 11, txt = 16$)

TABLE 6.2: The best Sen and Spe for all classifiers regardless of the ws and number of textons (txt) in the texton dictionary.

	$\boldsymbol{Sen}$ (%)	$\boldsymbol{Spe}$ (%)
ADTree	81.2 ± 10.8 ($ws = 9 \times 9, txt = 6$)	84.6 ± 12.7 ($ws = 9 \times 9, txt = 6$)
BNet	**86.7 ± 6.5** ($ws = 11 \times 11, txt = 16$)	**93.6 ± 6.7** ($ws = 11 \times 11, txt = 6$)
C4.5	86.2 ± 8.9 ($ws = 11 \times 11, txt = 6$)	87.5 ± 11.2 ($ws = 11 \times 11, txt = 6$)
k-NN	84.7 ± 8.2 ($ws = 9 \times 9, txt = 6$)	88.3 ± 9.4 ($ws = 11 \times 11, txt = 6$)
MLP	85.8 ± 10.4 ($ws = 11 \times 11, txt = 16$)	84.8 ± 10.8 ($ws = 11 \times 11, txt = 16$)
NB	83.8 ± 10.9 ($ws = 11 \times 11, txt = 16$)	83.6 ± 15.4 ($ws = 11 \times 11, txt = 16$)
RF	86.3 ± 7.6 ($ws = 11 \times 11, txt = 6$)	92.6 ± 8.3 ($ws = 11 \times 11, txt = 6$)
SL	78.6 ± 12.6 ($ws = 9 \times 9, txt = 16$)	82.1 ± 15.2 ($ws = 9 \times 9, txt = 16$)

6.6.1 Results for all Classifiers

This section presents the results for all classifiers based on the the ws and txt parameters. One of the main challenges in developing a texton based approach in texture classification is finding the best ws and txt as these parameters can influence the final results. For this purpose, we used the following $ws = 3\times3, 5\times5, 7\times7, 9\times9, 11\times11$ and 13×13 (top row). On the other hand, we tested the following $txt = 6, 10, 12, 16, 20, 24$ and 30 (left column). Note that these numbers represent the number of textons for both classes (e.g. $txt = 6$, 3 textons for benign and 3 textons for malignant class).

Tables 6.4 and 6.5, shows A_z and CA results, respectively using ADTree classifier

TABLE 6.3: The p values between the best A_z and CA against the other results regardless of the ws and txt.

	$\boldsymbol{A_z}$	$\boldsymbol{CA}$	***Sen***	***Spe***
ADTree	p <0.0001	p=0.00013	p=0.0034	p <0.0001
BNet	-	-	-	-
C4.5	p=0.0016	p=0.004	p=0.7642	p=0.0017
k-NN	p <0.0001	p=0.025	p=0.2005	p=0.0021
MLP	p <0.0001	p=0.0001	p=0.6241	p <0.0001
NB	p <0.0001	p=0.0038	p=0.126	p <0.0001
RF	p =0.016	p=0.095	p=0.7872	p=0.5287
SL	p <0.0001	p=0.0004	p=0.0001	p <0.0001

which produced best A_z=83.2% and CA=75.3%. At the smaller ws, the classifiers produced A_z ranging from 71% to 79% with slightly better at $ws = 7 \times 7$ compared to $ws = 3 \times 3$. It produced the best A_z at 9×9 (A_z>80% regardless to the number of textons).

The results for the best classifier in the proposed methods can be found in Tables 6.6 and 6.7. The BNet classifier performed best when features were extracted using a larger ws (e.g. 11×11). At $ws = 11 \times 11$ and $txt = 16$ the BNet outperformed the other predictive models with A_z=92.8% and CA=84%. At $ws = 13 \times 13$ the BNet predictive model decreased 1.5% and on average it can also be seen that all A_z and CA went down regardless to the number of textons. One noticeable pattern can be seen from the results in Tables 6.6 and 6.7 is at $ws = 3 \times 3$ until 11×11 the the

TABLE 6.4: A_z (%) values for ADTree classifier.

	$\boldsymbol{3 \times 3}$	$\boldsymbol{5 \times 5}$	$\boldsymbol{7 \times 7}$	$\boldsymbol{9 \times 9}$	$\boldsymbol{11 \times 11}$	$\boldsymbol{13 \times 13}$
6	74.6 ± 11.9	77.3 ± 12.1	79.4 ± 12.4	**83.8 ± 10.8**	83.2 ± 10.6	82.3 ± 10.7
10	74.3 ± 11.1	76.5 ± 11.9	77.6 ± 13.0	82.9 ± 9.8	80.7 ± 10.7	80.2 ± 9.9
12	74.5 ± 11.2	75.9 ± 11.9	77.3 ± 12.2	81.9 ± 9.7	80.2 ± 11.5	80.0 ± 10.7
16	73.8 ± 11.5	76.2 ± 11.7	77.0 ± 12.0	82.1 ± 10.7	82.5 ± 9.6	81.3 ± 10.2
20	73.1 ± 11.2	75.1 ± 11.3	76.1 ± 10.1	81.6 ± 10.2	76.5 ± 12.9	76.3 ± 11.9
24	72.6 ± 11.6	75.6 ± 11.9	75.6 ± 11.8	80.9 ± 9.9	76.6 ± 12.1	75.9 ± 11.5
30	72.3 ± 11.1	74.3 ± 12.0	75.1 ± 12.1	80.7 ± 10.9	76.3 ± 12.6	75.5 ± 11.7

range of A_z value increased around 4% to 5% at each ws and gradually decreased at 13×13. In terms of the number of textons, it did not effect the performance much on both metrics.

In comparison to the previous method in Chapter 5, the C4.5 classifier performed very well in this method with A_z=87.9% and CA=78.6% as shown in Tables 6.8 and 6.9, respectively. The A_z value gradually increased from the smallest ws until 11×11 which is similar to the previous classifiers. Nevertheless, there is no clear pattern of changes in terms of accuracy as most CA values are between the range of 72% to 75% regardless of ws and txt.

Another classifier which performed quite well in comparison to the features extracted in the previous chapter is k-NN with A_z=86.7% and CA=80.2% as shown in Table 6.10 and 6.11, respectively. In terms of area under the curve, the classifier performed better with smaller number of textons regardless of the ws. This can be seen in Table 6.10 when most A_z values are above 80% at $txt = 6$. As the txt value increases the A_z value decreases to around 70%. The lowest accuracy was produced at the largest ws of 13×13 with 30 textons in the texton dictionary.

Tables 6.12 and 6.13 show the results of MLP classifiers. Regardless of ws and txt, the results show that the MLP classifier performed 1.4% and 7.9% worst than its previous A_z and CA, respectively in the previous chapter. For instance, the best CA=76.2% at $ws = 11 \times 11$ and $txt = 16$, whereas in the previous chapter its best A_z=84.1% at 15×15 and on the same ws it produced 81%. Nevertheless, regardless of the number of textons on average its best A_z=84.4% at 9×9.

TABLE 6.5: CA (%) values for ADTree classifier.

	3 × 3	**5 × 5**	**7 × 7**	**9 × 9**	**11 × 11**	**13 × 13**
6	70.4 ± 18.2	71.0 ± 13.3	70.3 ± 14.6	71.9 ± 15.1	72.7 ± 14.6	70.9 ± 15.1
10	73.0 ± 18.3	69.5 ± 15.6	68.2 ± 14.4	69.7 ± 16.7	70.7 ± 16.3	70.1 ± 15.5
12	72.5 ± 18.6	70.8 ± 15.4	67.4 ± 14.9	69.3 ± 12.8	70.5 ± 16.3	70.3 ± 16.1
16	72.4 ± 19.2	68.6 ± 16.1	66.9 ± 15.4	69.3 ± 10.7	**75.3 ± 13.6**	74.3 ± 12.9
20	72.4 ± 19.9	68.3 ± 16.2	68.5 ± 14.5	69.2 ± 14.9	69.8 ± 16.2	68.3 ± 15.3
24	72.1 ± 20.7	68.1 ± 17.5	65.8 ± 16.3	70.0 ± 14.9	68.4 ± 16.8	67.1 ± 16.1
30	70.6 ± 21.7	63.9 ± 21.1	65.3 ± 15.1	69.1 ± 15.1	69.7 ± 16.8	68.1 ± 15.3

TABLE 6.6: A_z (%) values for BNet classifier.

	3 × 3	**5 × 5**	**7 × 7**	**9 × 9**	**11 × 11**	**13 × 13**
6	74.4 ± 11.9	79.2 ± 11.5	83.7 ± 10.7	89.5 ± 7.9	90.8 ± 7.8	89.4 ± 7.7
10	75.4 ± 11.5	80.1 ± 11.4	84.3 ± 10.2	90.3 ± 6.7	92.0 ± 6.9	91.3 ± 7.5
12	75.5 ± 11.5	80.2 ± 11.0	84.8 ± 10.1	90.0 ± 7.1	91.6 ± 6.1	90.7 ± 6.9
16	75.4 ± 11.5	81.7 ± 10.6	85.0 ± 9.9	90.9 ± 7.1	**92.8 ± 5.9**	91.3 ± 5.7
20	75.5 ± 11.4	80.5 ± 11.2	85.3 ± 9.7	90.8 ± 7.0	91.4 ± 6.0	90.7 ± 5.9
24	75.3 ± 11.4	80.9 ± 11.5	85.4 ± 9.6	90.8 ± 6.8	91.5 ± 6.3	90.1 ± 6.1
30	76.3 ± 10.8	81.0 ± 10.9	84.2 ± 8.9	90.9 ± 7.3	91.8 ± 5.9	91.1 ± 6.0

TABLE 6.7: CA (%) values for BNet classifier.

	3 × 3	**5 × 5**	**7 × 7**	**9 × 9**	**11 × 11**	**13 × 13**
6	67.2 ± 15.8	70.0 ± 15.2	73.6 ± 13.2	78.6 ± 10.2	80.5 ± 10.8	78.1 ± 10.2
10	69.1 ± 13.2	70.4 ± 14.1	74.0 ± 12.4	80.1 ± 9.9	83.0 ± 8.5	82.3 ± 9.5
12	68.1 ± 13.3	70.4 ± 14.5	74.5 ± 11.7	80.3 ± 10.1	82.1 ± 10.1	81.7 ± 10.2
16	67.4 ± 14.1	72.3 ± 14.2	75.1 ± 11.8	81.4 ± 8.9	**84.0 ± 7.0**	82.8 ± 7.5
20	67.2 ± 14	70.2 ± 14.8	75.7 ± 12.6	81.9 ± 8.0	82.0 ± 8.9	81.5 ± 8.3
24	67.5 ± 13.7	71.6 ± 14.6	75.6 ± 12.1	81.4 ± 8.9	83.0 ± 8.3	82.3 ± 8.9
30	68.0 ± 12.7	70.8 ± 14.3	75.3 ± 11.9	81.8 ± 8.2	82.8 ± 8.7	81.9 ± 7.9

TABLE 6.8: A_z (%) values for C4.5 classifier.

	3 × 3	**5 × 5**	**7 × 7**	**9 × 9**	**11 × 11**	**13 × 13**
6	71.6 ± 10.8	78.7 ± 11.6	82.4 ± 11.5	87.6 ± 8.7	**87.9 ± 8.6**	86.2 ± 8.3
10	74.4 ± 11.2	80.0 ± 11.1	83.1 ± 10.1	87.1 ± 7.9	85.2 ± 8.5	84.1 ± 8.1
12	75.5 ± 10.6	80.3 ± 10.7	82.8 ± 10.2	85.9 ± 7.7	83.3 ± 8.4	82.3 ± 7.9
16	76.1 ± 10.7	80.9 ± 11.1	81.2 ± 10.0	84.3 ± 7.9	82.4 ± 8.4	81.1 ± 8.3
20	75.4 ± 10.2	79.7 ± 10.7	80.0 ± 9.8	81.8 ± 7.4	76.9 ± 8.1	75.7 ± 7.7
24	75.4 ± 10.4	79.5 ± 10.8	77.9 ± 9.3	80.5 ± 7.2	75.7 ± 8.9	74.5 ± 8.8
30	70.2 ± 9.8	78.1 ± 10.2	77.8 ± 9.1	77.6 ± 6.9	74.2 ± 9.2	73.2 ± 9.3

The NB classifier is among the simplest and fastest classifier employed in the proposed method. The results shown in Tables 6.14 and 6.15 indicate that the NB classifier produced an acceptable performance with the best A_z=82.5% and CA=76.7%. In

TABLE 6.9: CA (%) values for C4.5 classifier.

	3 × 3	**5 × 5**	**7 × 7**	**9 × 9**	**11 × 11**	**13 × 13**
6	71.2 ± 17.6	74.4 ± 11.5	74.2 ± 11.9	76.8 ± 10.6	**78.6 ± 10.5**	75.2 ± 10.5
10	74.7 ± 15.5	74.6 ± 12.2	74.8 ± 9.3	76.3 ± 10.1	76.5 ± 11.3	74.3 ± 11.6
12	74.6 ± 15.2	75.8 ± 11.2	74.8 ± 9.7	75.4 ± 10.4	75.7 ± 11.7	73.3 ± 11.1
16	74.5 ± 14.8	75.4 ± 11.4	73.7 ± 9.4	75.4 ± 9.9	77.7 ± 10.4	75.7 ± 10.1
20	74.5 ± 15.2	74.8 ± 10.3	73.2 ± 8.9	74.5 ± 9.1	72.5 ± 10.7	70.5 ± 10.8
24	74.5 ± 14.9	74.0 ± 10.6	72.1 ± 8.9	74.0 ± 9.2	72.0 ± 10.7	71.3 ± 10.5
30	73.6 ± 15.1	73.4 ± 10.3	71.3 ± 9.3	73.0 ± 8.6	71.2 ± 10.6	70.6 ± 10.1

TABLE 6.10: A_z (%) values for k-NN classifier.

	3 × 3	**5 × 5**	**7 × 7**	**9 × 9**	**11 × 11**	**13 × 13**
6	76.3 ± 11.3	80.2 ± 11.3	82.7 ± 9.9	**86.9 ± 7.5**	85.2 ± 7.1	85.6 ± 7.4
10	77.3 ± 10.6	80.0 ± 10.1	80.8 ± 9.1	83.5 ± 6.8	80.1 ± 7.9	79.1 ± 7.9
12	77.5 ± 10.7	79.4 ± 9.6	78.3 ± 8.7	80.4 ± 7.5	79.0 ± 8.5	78.8 ± 8.3
16	77.4 ± 10.5	77.0 ± 9.4	74.8 ± 8.4	77.3 ± 7.4	77.7 ± 8.2	76.6 ± 8.1
20	76.4 ± 10.2	73.6 ± 8.6	72.2 ± 8.4	75.5 ± 7.4	72.8 ± 8.4	74.0 ± 8.9
24	76.0 ± 10.2	71.7 ± 8.3	70.1 ± 7.8	74.5 ± 7.9	71.6 ± 9.2	69.9 ± 9.5
30	75.1 ± 9.4	69.8 ± 8.1	69.8 ± 7.3	73.1 ± 7.4	70.7 ± 9.1	69.8 ± 9.3

TABLE 6.11: CA (%) values for k-NN classifier.

	3 × 3	**5 × 5**	**7 × 7**	**9 × 9**	**11 × 11**	**13 × 13**
6	71.2 ± 17.6	74.8 ± 11.3	74.3 ± 11.2	78.1 ± 9.6	**80.2 ± 9.1**	78.9 ± 9.3
10	74.9 ± 15.4	74.7 ± 11.7	74.2 ± 9.1	75.9 ± 8.5	74.3 ± 9.9	76.3 ± 9.4
12	75.0 ± 14.9	75.5 ± 10.8	73.2 ± 8.7	74.3 ± 8.5	73.6 ± 10.1	72.1 ± 10.3
16	74.6 ± 14.5	73.4 ± 10.3	71.1 ± 8.5	73.1 ± 8.3	74.3 ± 9.6	73.2 ± 9.8
20	74.5 ± 14.7	72.2 ± 9.5	69.6 ± 8.2	71.8 ± 8.0	70.0 ± 9.6	69.3 ± 9.7
24	74.5 ± 14.4	71.1 ± 9.6	68.7 ± 8.4	71.4 ± 8.4	68.9 ± 10.1	68.5 ± 10.4
30	73.3 ± 13.9	69.9 ± 9.3	67.8 ± 8.5	70.6 ± 8.2	68.5 ± 10.5	67.7 ± 10.6

terms of accuracy, results using small ws (e.g. 3×3 and 5×5) indicate poor performance of below 65%. It produced good results when employed in the previous method (Chapter 5) with A_z=87.3% and 86.9% at $ws = 9 \times 9$ and 11×11, respectively. On

TABLE 6.12: A_z (%) values for MLP classifier.

	3 × 3	**5 × 5**	**7 × 7**	**9 × 9**	**11 × 11**	**13 × 13**
6	73.1 ± 12.4	70.5 ± 16.2	79.8 ± 12.3	85.1 ± 10.7	84.0 ± 10.3	81.3 ± 10.9
10	74.8 ± 11.4	77.8 ± 11.9	79.1 ± 12.3	85.0 ± 9.1	83.1 ± 11.0	80.9 ± 11.3
12	75.0 ± 11.5	77.5 ± 11.8	80.0 ± 11.4	84.4 ± 9.2	83.1 ± 10.0	80.5 ± 11.1
16	75.3 ± 11.1	78.1 ± 12.2	79.1 ± 13.1	84.5 ± 9.4	**85.6 ± 9.3**	82.2 ± 9.8
20	75.7 ± 11.7	77.2 ± 12.2	78.0 ± 12.2	84.3 ± 9.2	80.5 ± 10.9	79.8 ± 10.8
24	76.1 ± 12.2	77.7 ± 12.4	78.2 ±11.8	84.1 ± 9.3	80.5 ± 11.4	79.3 ± 11.3
30	75.9 ± 11.9	77.5 ± 11.4	78.3 ± 11.2	83.7 ± 9.5	79.8 ± 11.8	78.5 ± 12.0

TABLE 6.13: CA (%) values for MLP classifier.

	3 × 3	**5 × 5**	**7 × 7**	**9 × 9**	**11 × 11**	**13 × 13**
6	70.5 ± 21.4	75.8 ± 13.6	70.7 ± 15.1	72.4 ± 13.3	72.8 ± 14.8	70.2 ± 14.2
10	71.6 ± 19.6	71.0 ± 14.5	67.6 ± 15.1	73.4 ± 11.6	71.5 ± 14.6	69.3 ± 15.7
12	72.2 ± 18.4	71.6 ± 13.8	70.0 ± 13.3	71.3 ± 11.7	71.4 ± 14.8	69.8 ± 15.5
16	71.8 ± 18.5	70.8 ± 13.5	69.7 ± 13.1	71.5 ± 12.5	**76.2 ± 11.5**	72.0 ± 12.2
20	71.5 ± 18.1	71.2 ± 12.9	68.6 ± 12.9	71.6 ± 11.1	71.8 ± 12.7	68.8 ± 13.9
24	70.9 ± 19.1	70.7 ± 13.5	68.3 ± 13.7	72.3 ± 11.1	71.0 ± 13.3	68.3 ± 13.7
30	70.5 ± 18.8	70.6 ± 13.3	68.5 ± 12.3	72.1 ± 10.9	70.2 ± 12.7	67.8 ± 14.2

average, using the same scales the NB classifier produced A_z=81.5% and 78% which is statistically significant ($p < 0.01$).

Tables 6.16 and 6.17 show the results of the second best classifier in our experiments which is the RF. In terms of A_z, the RF performed best at larger scales (e.g. 9×9 and 11×11) with $txt = 6$ or 10. In our experiment, using the maximum number of texton ($txt = 30$) decreased the A_z from 89.5% to 81.6%, which was statistically significant ($p < 0.001$). However, at $ws = 3 \times 3$ the A_z values show a very small variation ($\sigma = 0.53$) regardless of txt. On the other hand, both metrics are highest at $ws = 11 \times 11$ and lowest at $ws = 5 \times 5$.

Finally, Tables 6.18 and 6.19 show the results of SL classifier. The results show an acceptable performance with best A_z=81.4% and CA=75.7%. On average, the SL classifier better A_z at $ws = 9 \times 9$ but produced better CA at $ws = 11 \times 11$. Interestingly, in terms of accuracy, it has better performance on average at $ws = 3 \times 3$

TABLE 6.14: A_z (%) values for NB classifier.

	3 × 3	**5 × 5**	**7 × 7**	**9 × 9**	**11 × 11**	**13 × 13**
6	74.3 ± 12.4	75.8 ± 13.6	76.6 ± 14.8	82.2 ± 12.3	80.0 ± 12.7	79.9 ± 12.3
10	74.9 ± 11.6	76.7 ± 12.7	76.4 ± 13.8	82.0 ± 10.6	79.1 ± 12.5	78.5 ± 12.8
12	75.3 ± 11.7	76.3 ± 12.3	76.7 ± 13.8	81.6 ± 11.1	**82.5 ± 11.1**	80.1 ± 11.3
16	74.6 ± 11.8	75.8 ± 12.7	76.7 ± 13.1	81.3 ± 11.1	74.8 ± 19.8	75.3 ± 20.1
20	74.6 ± 11.8	75.8 ± 12.7	76.7 ± 13.1	81.3 ± 11.1	74.8 ± 19.8	75.3 ± 20.1
24	74.5 ± 11.4	76.4 ± 12.4	76.1 ± 13.0	80.8 ± 10.9	75.3 ± 14.9	74.3 ± 15.3
30	74.9 ± 10.5	75.8 ± 11.8	75.7 ± 12.7	81.5 ± 11.4	75.7 ± 13.6	73.2 ± 14.6

TABLE 6.15: CA (%) values for NB classifier.

	3 × 3	**5 × 5**	**7 × 7**	**9 × 9**	**11 × 11**	**13 × 13**
6	62.8 ± 17.9	63.2 ± 18.6	64.1 ± 20.0	65.3 ± 17.8	70.0 ± 19.6	67.3 ± 19.2
10	60.0 ± 18.6	62.7 ± 19.3	64.5 ± 20.0	68.7 ± 13.9	71.1 ± 19.4	70.2 ± 19.6
12	60.1 ± 18.8	61.5 ± 19.6	64.4 ± 19.5	69.3 ± 13.0	71.6 ± 18.9	70.5 ± 18.2
16	59.9 ± 18.9	64.2 ± 18.7	64.4 ± 19.5	70.3 ± 17.5	**76.7 ± 15.4**	71.1 ± 15.3
20	60.2 ± 18.7	61.6 ± 19.4	64.8 ± 20.0	69.4 ± 17.4	69.0 ± 19.8	67.5 ± 17.5
24	61.0 ± 18.5	64.2 ± 18.9	64.7 ± 19.5	69.6 ± 17.2	69.3 ± 19.9	67.3 ± 18.1
30	61.5 ± 18.3	63.1 ± 19.0	64.5 ± 19.3	68.7 ± 18.2	70.2 ± 19.1	68.1 ± 19.2

TABLE 6.16: A_z (%) values for RF classifier.

	3 × 3	**5 × 5**	**7 × 7**	**9 × 9**	**11 × 11**	**13 × 13**
6	76.3 ± 11.3	80.3 ± 11.3	83.6 ± 9.8	88.2 ± 7.4	**89.5 ± 7.1**	86.2 ± 8.1
10	77.3 ± 10.5	80.6 ± 9.9	83.0 ± 8.8	87.2 ± 6.6	86.7 ± 7.8	85.5 ± 8.2
12	77.6 ± 10.6	80.4 ± 9.6	81.9 ± 8.7	85.7 ± 6.7	86.0 ± 8.0	82.7 ± 7.9
16	77.5 ± 10.5	79.5 ± 9.3	80.4 ± 8.7	85.1 ± 7.5	86.3 ± 8.1	83.6 ± 8.2
20	76.9 ± 10.2	77.3 ± 9.1	79.3 ± 8.9	84.4 ± 7.4	82.6 ± 8.9	81.1 ± 8.2
24	76.6 ± 10.1	76.9 ± 8.9	78.8 ± 9.1	83.9 ± 7.6	82.0 ± 9.4	81.3 ± 9.5
30	76.2 ± 9.4	75.9 ± 9.1	78.3 ± 9.9	83.4 ± 8.4	81.6 ± 9.7	80.9 ± 8.9

TABLE 6.17: CA (%) values for RF classifier.

	3 × 3	**5 × 5**	**7 × 7**	**9 × 9**	**11 × 11**	**13 × 13**
6	71.2 ± 17.5	74.9 ± 11.3	74.6 ± 10.9	78.4 ± 9.6	**81.1 ± 9.3**	77.8 ± 9.8
10	74.8 ± 15.4	74.8 ±11.7	75.2 ± 8.9	77.7 ± 8.7	77.5 ± 9.9	75.3 ± 9.5
12	75.0 ± 14.9	75.9 ± 10.6	74.9 ± 8.9	76.9 ± 8.7	77.8 ± 9.0	74.3 ± 8.9
16	74.6 ± 14.5	74.5 ± 10.1	73.9 ± 8.8	76.3 ± 8.7	77.0 ± 10.0	75.1 ± 9.5
20	74.6 ± 14.6	74.0 ± 9.4	73.4 ± 8.6	75.3 ± 8.2	74.0 ± 9.7	73.0 ± 9.3
24	74.7 ± 14.2	73.6 ± 9.7	72.7 ± 9.1	75.2 ± 8.4	73.0 ± 10.2	74.0 ± 10.3
30	73.6 ± 13.6	73.1 ± 9.7	72.5 ± 9.3	74.3 ± 8.6	72.1 ± 11.0	72.2 ± 10.8

TABLE 6.18: A_z (%) values for SL classifier.

	3 × 3	**5 × 5**	**7 × 7**	**9 × 9**	**11 × 11**	**13 × 13**
6	53.1 ± 6.2	71.7 ± 14.2	72.6 ± 16.0	80.5 ± 10.3	76.9 ± 14.6	74.3 ± 14.9
10	70.3 ± 13.4	74.1 ± 13.3	74.7 ± 14.0	80.6 ± 11.7	76.9 ± 12.9	74.1 ± 14.3
12	71.8 ± 11.9	74.2 ± 12.4	75.6 ± 13.5	79.8 ± 10.8	77.1 ± 12.8	76.1 ± 12.3
16	72.6 ± 11.5	75.8 ± 12.3	76.1 ± 12.9	80.6 ± 11.2	81.2 ± 11.5	80.2 ± 11.3
20	72.8 ± 11.7	75.1 ± 12.1	75.8 ± 12.9	80.9 ± 10.9	74.5 ± 14.6	73.2 ± 15.1
24	73.3 ± 11.5	75.8 ± 12.6	75.6 ± 12.8	**81.4 ± 10.6**	75.8 ± 14.4	74.1 ± 14.1
30	74.3 ± 11.1	75.5 ± 12.1	75.3 ± 12.9	80.9 ± 11.7	75.6 ± 13.3	73.4 ± 12.9

TABLE 6.19: CA (%) values for SL classifier.

	3 × 3	**5 × 5**	**7 × 7**	**9 × 9**	**11 × 11**	**13 × 13**
6	71.7 ± 20.4	69.9 ± 16.8	65.7 ± 16.1	68.2 ± 16.3	67.2 ± 18.1	66.1 ± 19.1
10	71.0 ± 22.2	69.4 ± 16.1	64.2 ± 15.3	67.7 ± 15.6	70.6 ± 19.1	69.3 ± 19.3
12	71.9 ± 19.1	70.1 ± 13.6	66.5 ± 12.5	67.3 ± 14.6	71.0 ± 17.8	68.3 ± 18.2
16	72.0 ± 18.7	69.3 ± 14.6	66.9 ± 13.2	68.0 ± 15.0	**75.7 ± 14.2**	71.5 ± 14.5
20	72.1 ± 18.7	68.9 ± 13.9	66.2 ± 12.0	68.8 ± 14.2	69.0 ± 18.2	67.8 ± 18.3
24	71.8 ± 18.6	68.3 ± 14.7	65.8 ± 12.6	68.4 ± 13.7	68.3 ± 18.5	67.6 ± 18.7
30	71.4 ± 18.7	68.2 ± 14.1	64.3 ± 14.3	69.5 ± 13.6	69.0 ± 17.2	68.1 ± 17.8

in comparison to the other scales. However, in terms of A_z all results show around 80% regardless to txt at $ws = 9 \times 9$.

6.7 Parameter Optimisation

Similar in Chapter 5, we tested the performances of three classifiers ($k-$NN, ADTree and RF) when the parameters of the classifiers are varied. The same parameter ranges were used as mentioned in 5.6.

Figure 6.11 shows results for A_z and CA when k is varied from 1 to 41. The data was extracted using $ws = 11 \times 11$ and $txt = 12$ (6 texton per class). In terms of classification accuracy no significant difference was noticed as all CA were between 72% to 74%. Nevertheless, there was a significant difference in terms of A_z at $k = 1$ and $k = 5$. The A_z increased to just below 84% as the k increases in comparison to $k = 1$ where the A_z value is around 79%.

On the other hand, Figure 6.12 shows the results for the ADTree classifier using 22 different nB values. The classifier produced $A_z \leq 80\%$ and $CA \leq 70\%$ at $nB \leq 9$. At the default nB given in WEKA [2] produced around A_z=83% and CA=73%. At $nB \geq$

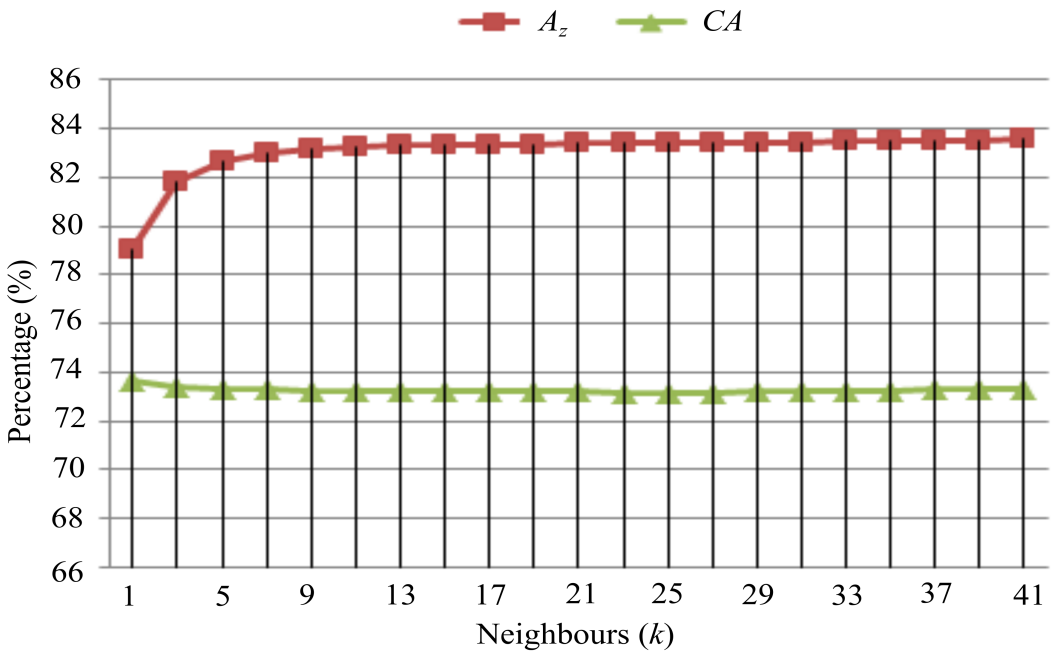

FIGURE 6.11: The A_z and CA values using different k values for the k-NN classifier. Default $k = 1$ in WEKA.

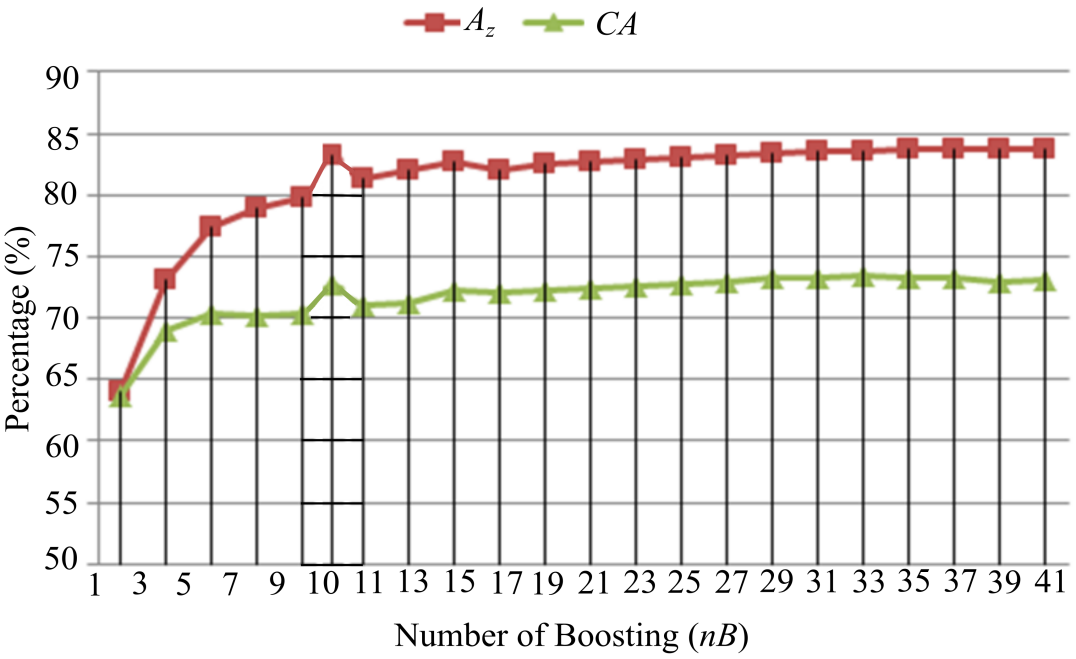

FIGURE 6.12: The A_z and CA values using different nB values for the ADTree classifier. Default $nB = 10$ in WEKA.

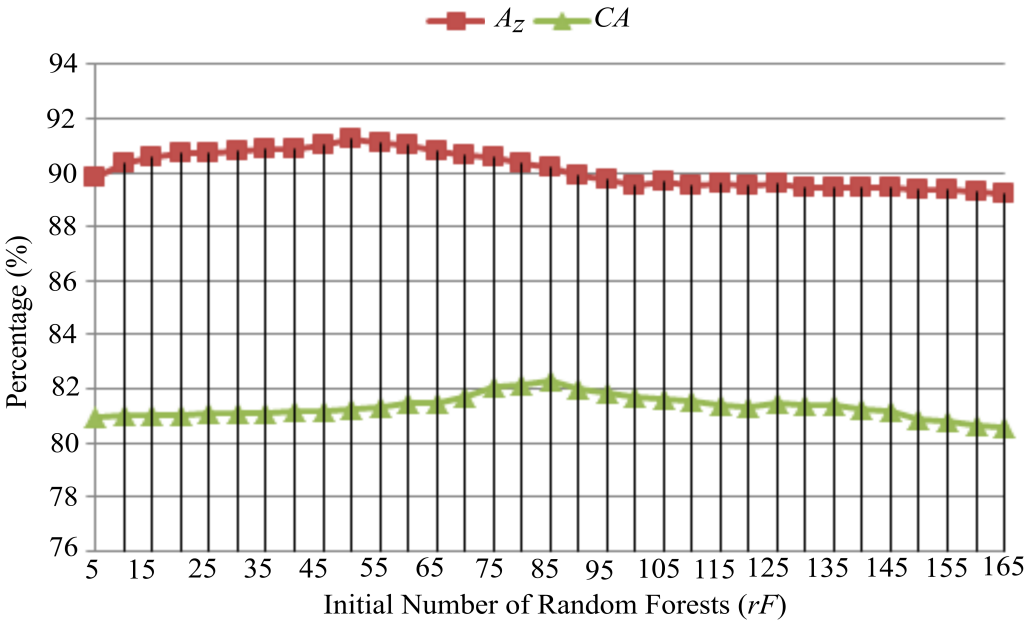

FIGURE 6.13: The A_z and CA values using different rF values for the RF classifier. Default $rF = 100$ in WEKA.

11 both metrics change around 1% before they gradually increase at $nB \geq 17$ and reaches $A_z \leq 83\%$ and $CA \leq 73\%$. As shown in Figure 6.12, the classifier produced the best A_z and CA at $nB = 10$. Our explaination for this behaviour are two-fold: first it may be caused that an optimal model for the data was achieved at $nB = 10$ which means adding more boosting iteration resulting an overfitting model and secondly adding more boosting iteration decreases the training error but increases the test error which affects the overall accuracy. In an early study conducted by the original authors, Freund and Manson [306] showed that a significant test error was encountered at $nB > 10$. In Figure 6.13, the RF classifier achieved the highest A_z=91% at $rF = 50$ with A_z just above 82% with $rF = 85$. Overall, there is no significance difference for both performance metrics using $rF = 5$ to 165, the CA=80%-82% and A_z=89%-91%. In this section, one visible pattern for all the classification results is after an optimal model (or optimal parameter) is achieved both metrics showed very little change in percentage (almost no difference) which may be caused by the size of our data (more than 150,000 instances). For example 100 misclassified instances has very little effect on the percentage.

6.8 Discussion

In our experiments, the results suggest that the number textons in the dictionary has a significant effect on both A_z and CA. For example ADTree, C4.5, k-NN and RF classifiers produced better results with smaller number of txt value. In contrast, the BNet and MLP classifiers performed better with $txt = 16$ or 20, whereas the SL and NB classifiers were not affected much by the number of texton in the dictionary. Both metrics are highly influenced by the ws used to construct the histograms (treated as feature vector) from the texton maps which is similar to the previous method proposed in Chapter 5. Furthermore, using the maximum value of ws (in our case 13×13), reduced the performance on both metrics. In terms of selecting the best ws and txt, most classifiers performed well at 9×9 and 11×11 with 6 or 16 textons (3 and 8 textons per class, respectively).

At a smaller ws, features are unable to characterise the actual texture (or in our case the texton distribution) as smaller ws tends to capture noise. On the other hand, at a larger ws (e.g. 9×9) features tend to be more reliable because 'noisy pixels' are shrunk due to dominating 'reliable pixels' (e.g. malignant pixels). However, when using a large ws (e.g. 13×13), the performance tends to decrease because the chance

of mixing up pixels from benign or malignant class is higher, hence altering the actual feature's representation of a particular class. On the other hand, the number of textons affects the complexity of the predictive model built by the classifier. For example using smaller number of textons yields to a less complex data (e.g. RF and ADTree) and reduces the amount of error in finding the boundary between two classes, hence increases the accuracy. In some classifiers using a slightly larger number of textons results a better classification models (e.g. BNet and MLP) due to the increase layers created in the model which lead to a better classification.

There are two main advantages of the proposed method: first it bypasses the typical feature extraction algorithm such filtering and convolution which can be computationally expensive and secondly it does not need the additional step of feature selection as the number of textons are already small (e.g. $txt = 6$ or 16). With a large number of features (e.g. in our previous method extracted 215 features), selecting the best features can be time consuming. Although feature selection can significantly improve classification results, it needs to be applied with a robust feature selection algorithm as discussed in the previous chapter and a study conducted by Niaf *et al.* [231]. The results suggested that even with $txt = 6$, our method can produce A_z>90% with the BNet classifier and using the simplest classifier (k-NN) still can achieve around 87%. In fact, all classifiers employed in this study produced A_z>80%.

Indeed, one of the limitations of this study is that we are unable to compare the results quantitatively with the existing methods due to absence of public datasets which is the major problem among the research communities in CAD-PC. Nevertheless, the results presented in this chapter further support the potential of CAD in detecting prostate cancer and minimising clinicians' deficiencies during the screening phase and hopefully can provide more reliable results. Moreover, we only performed three experiments (for three classifiers) in the parameter optimisation section and only used features extracted using $ws = 11 \times 11$.

6.9 Summary

To summarise this chapter, we have presented the first texton based approach for prostate cancer detection in the literature and experimental results show that all classifiers achieved A_z>80% regardless of ws and txt. The best A_z=92.8% was achieved

using the BNet classifier where the RF classifier produced $\approx 90\%$, which is similar to the recent method proposed in [46]. This chapter highlights the following points:

1. The proposed texton based approach performed better in comparison to the previous method on the basis of there is no need to combine several classifiers (e.g. Meta-Vote as previously used in Chapter 5) and there is no need to employ feature selection as the number of features is already minimised.

2. The proposed method produced better results with small number of textons which means our method tends to produce a less complex data and classification model hence decreases the training and testing time for reliable results.

3. This chapter further investigated the use of ws in CAD-PC study and found very similar results to the previous chapter (all classifiers achieved their peak performance at 9×9 or 11×11).

4. The experimental results show that the expensive feature extraction methods via filtering and convolution in CAD-PC could be bypassed by using a texton based approach.

In the next chapter, we will qualitatively compare our results with the state-of-the-art in the literature and discuss some of the existing methods and how our methods differ from their methods.

Chapter 7

Qualitative Comparisons

This chapter makes qualitative comparisons with the state-of-the-art methods described in the literature and covers the following: (a) why it is difficult to make quantitative comparisons in CAD-PC (b) a brief review some of the existing methods (c) qualitative comparisons against the actual clinical performance and (d) discussion about the performance results from different perspectives. Although it is difficult to draw a conclusion as to which method produced superior performance, experimental results indicate that the proposed methods in this thesis achieved similar results to the existing methods as well as human performance.

7.1 Introduction

Despite promising results of CAD-PC in assisting radiologists in diagnostic decision making, the major problem is the lack of publicly available datasets resulting in each group of researchers having their own different datasets. This makes quantitative comparisons difficult due to:

1. Differences in the type of modalities (different modalities such as T2-weighted (T2-W) MRI, diffusion (DWI) MRI, dynamic contrast enhanced (DCE) MRI, Magnetic resonance spectroscopy (MRS), etc.) and protocols used in the other studies. Many of the earlier studies were based on a single modality such as T1-W, T2-W and DCE [206, 207, 211, 346]. However, as the MRI technology evolved most of the later CAD-PC studies were based on multiparametric

MRI data, such as [46, 200, 226, 229, 231, 232, 311, 347]. This makes quantitative comparisons are difficult as the tissues' numerical representations are inconsistent in different modalities.

2. Absence of public datasets also makes quantitative comparisons among CADs in the literature difficult. Each team of researchers has their own datasets which cause a huge range of variability in terms of noise, acquisition protocol, the size of the dataset (e.g. number of patients) and image quality. For example the methods in [21, 211, 231] were evaluated based on datasets taken using 1.5 Tesla MRI scanners whereas the studies in [45, 46, 263] were evaluated on datasets taken from 3.0 Tesla MRI scanners. In terms of image clarity, the 3.0 Tesla MRI scanners produce better vascular imaging quality.

3. Studies were conducted within different regions of the prostate. For example, some studies [21, 192, 231, 236, 259] were conducted within the prostate PZ only and some [46, 64, 65, 229] took the whole prostate gland into account. Visually it is harder to detect and differentiate malignant regions within the CZ in comparison to the ones in the PZ. Although the study by Viswanath *et al.* [295] suggested that tumours in the CZ and PZ are significantly different, their dataset was acquired using 3.0 Tesla MRI scanner.

4. Finally, another difficulty is that the bases of evaluation have been different (e.g. volume, slice, regions or voxel/pixels). Pixel level evaluation is more challenging as the number of instances increases for a larger area of the region which makes the classifier model more complex. On the other hand, at region level evaluation the number of instances is limited to the number of regions annotated by the radiologists. For example the results reported in [21, 300] are based on 87 and 90 regions of interest (ROI), respectively annotated by the radiologists. Similarly the method in [211] was evaluated based on 130 selected ROI by the authors. This means the number of instances in the training and testing phases is quite small. In comparison, the number of instances in our study are around 150,000.

Nevertheless qualitative comparisons can roughly indicate the relative performance of the proposed methods in this thesis.

7.2 Comparisons

This section presents the CAD-PC results reported in the literature based on the current review conducted by Lemaître *et al.* [62] in 2015 which covers more than 35 methods from 2007 until 2014. The evaluation metrics used in most studies are area under the curve (A_z or AUC), accuracy (CA), sensitivity (Sen) and specificity (Spe). However, note that most of the authors did not report all metrics (indicated as '-'). The studies also covered both multiparametric and monoparametric CAD methods. Tables 7.1, 7.2 and 7.3 present the results of the existing methods in the literature ranked according to A_z, CA, Sen and Spe. The strength of the magnetic field used by the scanner for data acquisition is indicated as 1.5T or 3.0T.

TABLE 7.1: Qualitative comparison: Dynamic Contrast Enhanced (DCE), Diffusion Weighted (DWI), Magnetic Resonance Spectroscopy (MRS) and Magnetic Field Strength (MFS).

Authors	**Patients**	**Studied Zones**	**MRI Modality (MFS)**	A_z (%)	CA (%)	Sen (%)	Spe (%)
Vos *et al.* (2010) [21]	29	PZ only	T2-W + DCE (1.5T)	97	-	-	-
Lv *et al.* (2009) [211]	55	PZ only	T2-W (1.5T)	97	-	-	-
Peng *et al.* (2013) [45]	48	PZ and CZ	T2-W + DCE + DWI (3.0T)	95	-	82	95
Antic *et al.* (2013) [348]	53	PZ and CZ	T2-W + DWI (1.5T)	94	-	-	-
Lopes *et al.* (2011) [208]	17	PZ only	T2-W (1.5T)	93	-	93	93
Our method	**45**	**PZ only**	**T2-W (1.5T)**	**92.8**	**84**	**86.7**	**93.6**
Our method	**45**	**PZ only**	**T2-W (1.5T)**	**92.7**	**88.1**	**93.3**	**86.5**
Vos *et al.* (2008) [300]	34	PZ only	T2-W + DCE (1.5T)	92	-	-	-
Vos *et al.* (2008) [349]	29	PZ only	T2-W + DCE (1.5T)	91	-	-	-
Tiwari *et al.* (2013) [350]	19	PZ and CZ	T2-W + MRS (1.5T)	91	-	-	-
Ampeliotis *et al.* (2008) [207]	25	PZ only	T2-W + DCE (1.5T)	90	87	-	-
Tiwari *et al.* (2012) [65]	36	PZ and CZ	T2-W + MRS (1.5T)	90	-	-	-
Litjens *et al.* (2014) [46]	347	PZ and CZ	T2-W + DCE + DWI (3.0T)	89	-	-	-
Niaf *et al.* (2012) [231]	30	PZ only	T2-W+ DWI + DCE (1.5T)	89	-	-	-
Niaf *et al.* (2011) [351]	30	PZ only	T2-W+ DWI + DCE (1.5T)	89	-	-	-
Tiwari *et al.* (2013) [229]	29	PZ and CZ	T2-W + MRS (1.5T)	89	-	-	-
Giannini *et al.* (2011) [217]	10	PZ only	T2-W+ DWI + DCE (1.5T)	87	-	79	84

TABLE 7.2: Qualitative comparison (continues from Table 7.1).

Authors	**Patients**	**Studied Zones**	**MRI Modality (MFS)**	A_z (%)	CA (%)	Sen (%)	Spe (%)
Tiwari *et al.* (2009) [352]	15	PZ and CZ	T2-W + MRS (1.5T)	84	-	84	81
Chan *et al.* (2003) [233]	15	PZ only	T2-W + DWI (1.5T)	84	-	-	-
Liu *et al.* (2013) [194]	54	PZ and CZ	T2-W + DCE + DWI (3.0T)	83	-	-	-
Viswanath *et al.* (2009) [200]	6	PZ and CZ	T2-W + DCE (3.0T)	82	-	-	-
Puech *et al.* (2009) [346]	100	PZ and CZ	DCE (1.5T)	77	-	100	43
Viswanath *et al.* (2011) [64]	12	PZ and CZ	T2-W + DCE + DWI (3.0T)	77	-	-	-
Viswanath *et al.* (2012) [295]	22	PZ and CZ	T2-W (3.0T)	73	-	-	-
Langer *et al.* (2009) [353]	25	PZ only	T2-W + DCE + DWI (1.5T)	71	-	-	-
Liu *et al.* (2009) [234]	11	PZ only	T2-W+ DWI + DCE (1.5T)	-	89	90	88
Ampeliotis *et al.* (2007) [206]	25	PZ only	T2-W + DCE (1.5T)	-	84	-	-
Sung *et al.* (2011) [291]	42	PZ and CZ	DCE (3.0T)	-	83	77	77
Viswanath *et al.* (2008) [354]	16	PZ and CZ	T2-W (1.5T)	-	-	88	85
Tiwari *et al.* (2008) [355]	18	PZ and CZ	MRS (1.5T)	-	-	87	85
Our method	**37**	**PZ only**	**T2-W (1.5T)**	-	**86**	**87**	**86**
Parfait *et al.* (2012) [299]	22	PZ and CZ	MRS (3.0T)	-	-	84	97
Ozer *et al.* (2010) [198]	20	PZ only	T2-W+ DWI + DCE (1.5T)	-	-	78	74
Ozer *et al.* (2009) [197]	20	PZ only	T2-W+ DWI + DCE (1.5T)	-	-	76	75

TABLE 7.3: Qualitative comparison (continues from Table 7.2).

Authors	**Patients**	**Studied Zones**	**MRI Modality (MFS)**	A_z (%)	*CA* (%)	*Sen* (%)	*Spe* (%)
Artan *et al.* (2010) [196]	10	PZ only	T2-W+ DWI + DCE (1.5T)	-	78	74	82
Artan *et al.* (2009) [195]	21	PZ only	T2-W+ DWI + DCE (1.5T)	-	-	66	72
Matulewicz *et al.* (2014) [356]	18	PZ and CZ	MRS (3.0T)	-	-	63	99
Viswanath *et al.* (2008) [218]	6	PZ and CZ	T2-W + DCE (3.0T)	-	-	60	66

Results from all three tables show that the smallest and largest datasets were used by Viswanath *et al.* [218] and Litjens *et al.* [46] which covered 6 and 347 patients, respectively with an average around 37 patients per study. Only 10 studies were conducted based on datasets acquired using MRI scanner machine with MFS=3.0T. The number of CAD-PC studied within the PZ and the whole prostate gland are similar (20 and 18 studies conducted within the PZ and whole prostate gland, respectively including our proposed method). In terms of performance evaluation metrics more than 65% of the methods used A_z and approximately half of the methods used *Sen* and *Spe* to measure the performance of their methods. On the other hand, 67% of the methods are based on multiparametric MRI and only 33% are based on single modality. In terms of performance and the size of dataset, results in Tables 7.1, 7.2 and 7.3 indicate that there is no visible correlation between them. For example Viswanath *et al.* [200] achieved an A_z=82% based on 6 patients whereas Litjens *et al.* [46] achieved A_z=89% based on 347 patients. On the other hand, the method proposed by Vos *et al.* [21] was evaluated based on 29 patients achieved A_z=97% but in a larger dataset of 100 patients Puech *et al.* [346] reported an A_z=77%. Nevertheless, we do not know the number of instances used in the dataset. For example a large number of patients does not mean it has large number of instances (e.g. if the classification is based on ROI). In comparison, pixel based classification tend to have a larger number of instances although the number of patients is small. Unfortunately the majority of the studies did not report this information.

All methods in Table 7.1 achieved at least A_z=87%. The methods proposed by Vos *et al.* [21] and Lv *et al.* [211] achieved the highest A_z=97%. Vos *et al.* [21] proposed a method using features extracted from quantitative pharmacokinetic (PK) maps and T2-W MRI before training a SVM to calculate the malignancy likelihood of each lesion. However, the method was tested on a small dataset of 87 ROI taken from 29 patients. In an earlier study, using similar features Vos *et al.* [300] reported an A_z=92% based on 90 ROI taken from 34 patients. On the other hand, Lv *et al.* [211] used analysis of histogram fractal dimension (HFD) and texture fractal dimension (TFD) information on a single modality of T2-W MRI. Although the study covered 55 patients, the actual evaluation was based on selected 130 ROI of 12×12 pixels (which means only a small part of the PZ region was covered). In fact, Lv *et al.* [211] did not perform cross validation to further evaluate their method. In our study, we performed 9-FCV as well as tested the proposed method on 418 PZ regions.

Peng *et al.* [45] reported A_z=95% using three modalities of T2-W, DCE, and DWI-MRI and extracted 10^{th} percentile apparent diffusion coefficient (ADC), average ADC, and T2-W skewness. Subsequently, individual image features were combined using linear discriminant analysis (LDA) to perform leave-one-patient-out cross validation. From an evaluation point of view, their study is similar to the studies in [21, 211]. Although Peng *et al.* [45] reported that their study covered 48 patients, the actual evaluation was based on 104 ROI (61 malignant ROI, 43 normal ROI). In a smaller study of 17 patients, Lopes *et al.* [208] concluded that classical texture features (such as Haralick, wavelet, and Gabor filters) are less discriminant in classifying malignant and benign regions in comparison to fractal and multifractal features. In their study, they combined fractal and multifractal features and employed SVM and AdaBoost classifiers to get A_z=93% in comparison to the combination of classical texture features (A_z=88%). Antic *et al.* [348] achieved an A_z of 1% better (in comparison to the method of Lopes *et al.* [208]) by combining only T2-W and DWI with a larger number of patients (54) covered the whole prostate gland. In comparison, our proposed methods achieved similar results qualitatively with some of the methods in the literature regardless of the size of dataset, modality and studied zones.

In [300] the authors extracted statistical features such as percentile 25%, 75% and K^{trans} from each of the annotated ROI in DCE images before employing SVMs for classification. Vos *et al.* [349] reported $Az = 92\%$ for malignant and benign discrimination and 83% for suspicious malignant and benign discrimination which is similar to the earlier study. Ampeliotis *et al.* [206] who conducted two similar studies on the

same dataset reported an accuracy of 84% but by applying feature generation transform which reduces the dimension of the feature vectors the accuracy increased to 87% with A_z=90% in [207]. Niaf *et al.* [231] extracted 140 texture features from 180 ROI (30 patients) and achieved A_z=89% which is similar to the methods in [19, 25, 88]. Niaf *et al.* [231] compared the performance of four different classifiers (SVM, LDA, k-NN and NB) based on four different feature selection methods. Further, their results showed that employing feature selection significantly improved the performance of their method and gradient features showed a high discriminant capability in their study.

The methods in [46, 231, 353] attempted to cover the whole prostate gland. Recently, Litjens *et al.* [46] conducted a study which covered 347 patients and reported A_z=89%. Their method consisted of two stages: in the first stage the prostate gland was segmented using a multi-atlas-based segmentation method and features based on intensity, anatomical, pharmacokinetic, texture and blobness were calculated. Subsequently, each voxel was classified using GentleBoost and RF classifiers to generate a likelihood map. On each likelihood map local maxima detection was performed to capture ROI with the highest probability of being malignant. A method by Tiwari *et al.* [229] which is based on multi-kernel graph embedding in T2-W and MRS produced A_z=89% covering 29 patients. The method [229] was also based on a two-stage classification approach: in the first stage, a voxel based classification was performed by employing a random forest classifier in conjunction with the SeSMiK-GE based data representation and a probabilistic pairwise Markov Random Field (MRF) algorithm to identify malignant ROI. Subsequently, each of the segmented malignant ROI was classified as either high or low Gleason grade. Using the same method, in a smaller study of 19 patients Tiwari *et al.* [350] reported an A_z=91%. Later, Tiwari *et al.* [65] proposed a data integration framework for T2-W and MRS for prostate cancer detection. Texture descriptors such as Gabor, gradient, first- and second-order statistical features were extracted from T2-W and wavelet features were extracted from MRS images. Both feature sets were fused (via dimensionality reduction) using their proposed framework before employing a probabilistic boosting tree (PBT), SVM and RF classifiers. They reported an improvement of at least A_z=5% in comparison to the results without using the proposed data integration framework.

On the other hand, Giannini *et al.* [217] developed a method based on the features extracted from registered images of different MRI modalities. The Bayesian classifier was used to fuse all the extracted features into a probability map (the region with the

highest probability of being malignant is considered cancerous region). Preliminary results show that the proposed method achieved A_z=87%. A study by Viswanath *et al.* [295] attempted to differentiate the textural characteristics of malignant regions within the CZ and PZ. They extracted 110 texture descriptors and reported that Haralick's features (e.g sum entropy and difference average) achieved A_z=73% in differentiating cancer regions within the PZ, whereas Gabor features performed better within the CZ (A_z=73%). However, in a study of multiparametric MRI Viswanath *et al.* [64] reported an increase of at least 4%. In a smaller dataset Viswanath *et al.* [200] (6 patients) A_z=82% although in the earlier study [218] they reported only 60% accuracy using the same dataset. In contrast to the study conducted by Chan *et al.* [233], they introduced a multichannel statistical classifier and applied it for prostate cancer detection. The method achieved a reasonable A_z=84% considering their texture features are only based on image intensity obtained from T2-W and DWI. The proposed methods of Puech *et al.* [346] and Langer *et al.* [353] achieved A_z<80% covering 100 and 25 patients, respectively. Puech *et al.* [346] used a scoring algorithm approach as part of their method. The scoring algorithm was firstly developed based on median and maximum 'wash-in' and 'wash-out' slope values and they used these values to assign malignancy likelihood score for each of the ROI. Langer *et al.* [353] combined three MRI modalities via several image fusion techniques and achieved a reasonable A_z=71% based on 25 patients.

In unsupervised CAD methods, Liu *et al.* [234] reported 89%, 88% and 90% accuracy, sensitivity and specificity, respectively based on 11 patients covering the PZ only. In their method the authors proposed a new segmentation method using fuzzy Markov random fields (fuzzy MRFs) for the segmentation of multispectral MRI. Instead of treating the MRF and class parameters separately, the authors combined these together which produced more accurate segmentation results. In comparison, our unsupervised method achieved 86%, 87% and 86% accuracy, sensitivity and specificity, respectively based on 37 patients in the same zone of the prostate. On the other hand, Sung *et al.* [291] investigated the value of CAD-PC in different prostate zones using DCE (3.0T) based on 42 patients. Their study revealed that accuracy, sensitivity, and specificity were 83%, 77%, and 77%, respectively, in the entire prostate; 77%, 91%, and 64%, respectively, in the TZ; and 89%, 89%, and 89%, respectively, in the PZ.

Ozer *et al.* [197, 198] who developed two different methods showed that the combination of the pharmacokinetic parameters derived from DCE MRI with T2 MRI and

DWI increased the segmentation accuracy significantly but showed similar sensitivities and specificities using SVM and Relevance Vector Machines (RVM) classifiers. Similarly, Artan *et al.* [196] showed that multispectral MRI helps to increase the accuracy of prostate cancer localisation compared to single modality. The authors developed a segmentation method by combining conditional random fields (CRF) with a cost-sensitive framework which incorporates spatial information. Considering the features are only derived from pixel intensities of multispectral MR modality images, the method achieved 78%, 74% and 82% accuracy, sensitivity and specificity, respectively. Nevertheless in a larger dataset of 21 patients in [195] the author reported 66% and 72% sensitivity and specificity, respectively. Finally, Parfait *et al.* [299] achieved a high specificity of 97% but a lower sensitivity of 84% based on MRS (3.0T) data whereas Matulewicz *et al.* [356] on the same modality reported 99% and 63% specificity and sensitivity, respectively.

7.3 Graphical Representations of the State-of-the-art

This section presents the performance distributions in all four metrics reported in the studies discussed previously. In each of the performance distributions, we calculated each metric average and the number of patients. Subsequently, the average line is drawn in each graph to visualise which methods achieved above the average of the-state-of-the-art. This is indicated within the magenta region called 'above the average zone'. This zone covers methods with performance above the average. The average of each metric is calculated by finding the summation of a particular metric and dividing it by the number of methods which have reported the corresponding metric in their studies. Note that studies were based on the recent review conducted by Lemaître *et al.* [62].

Figure 7.1 shows the distribution of A_z values against the number of patients from studies in the literature (only those studies which have reported A_z). The graph indicates that there is no correlation between these two variables (number of patients and A_z). Both of our proposed methods in Chapters 5 and 6 can be seen in green and red colours, respectively. Most studies covered 30 to 50 patients achieved A_z>90% and with 100 patients Puech *et al.* [346] reported A_z=77%. The study of Litjens *et al.* [46] which covered 347 patients achieved A_z=89%. The yellow lines indicate

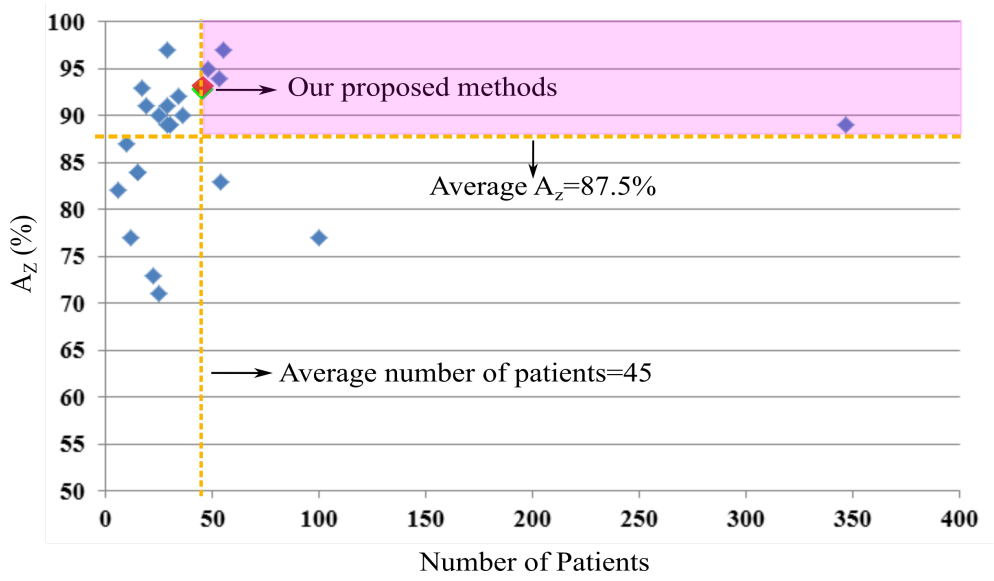

FIGURE 7.1: A scatter plot representation of the A_z against the number of patients.

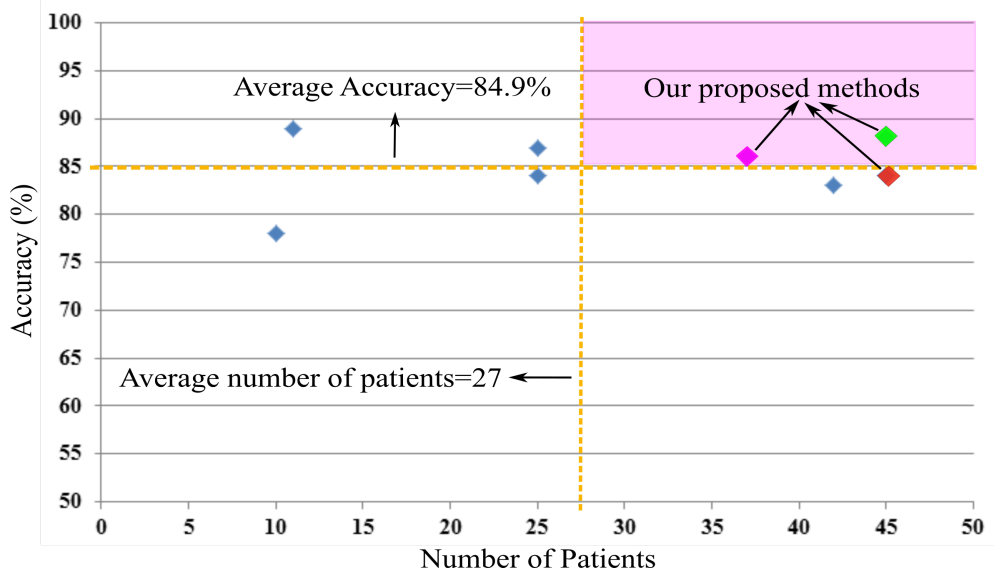

FIGURE 7.2: A scatter plot representation of the CA against the number of patients.

the average number of patients and average A_z based on 25 studies which have used this metric to measure the performance of their methods, resulting average number of patients is 45 and average A_z=87.5%. Based on these statistical figures, both of our proposed methods achieved above the average of the state-of-the-art.

In terms of accuracy, only eight studies reported the performance of their methods using this metric and only two of our proposed methods (both in Chapters 4 and 5) managed to reach the 'above the average zone' as shown in Figure 7.2. Although the studies in [206, 234] achieved CA=89% and 87%, respectively but their methods were evaluated based on 25 and 11 patients respectively which is below the average number

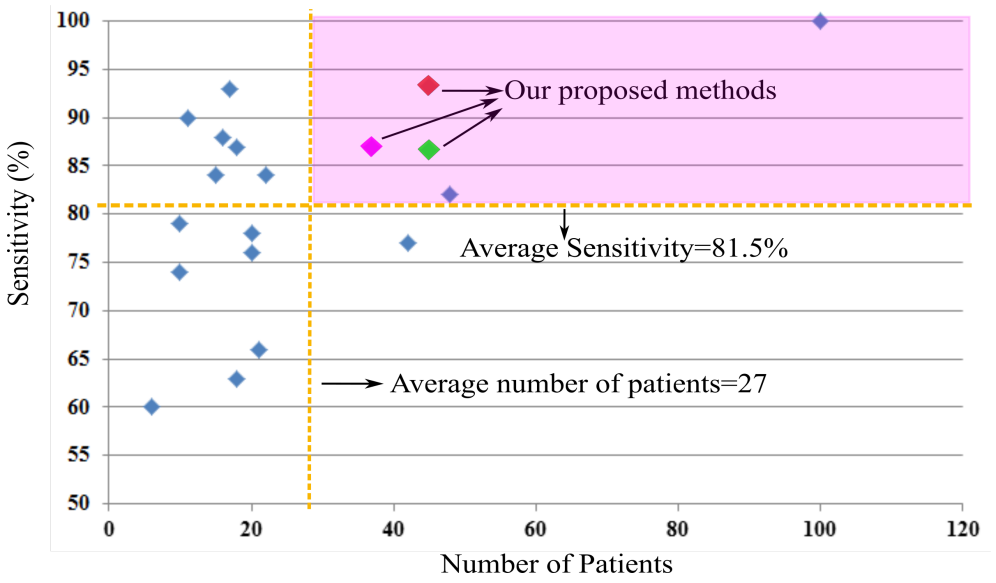

FIGURE 7.3: A scatter plot representation of the *Sen* against the number of patients.

of patients of the eight studies. On the other hand, in a larger study of 42 patients based on a single modality (DCE (3.0T)) Sung *et al.* [291] reported CA=83%.

On the other hand, Figure 7.3 shows the distribution of *Sen* against the number of patients (only those studies have reported *Sen*). Lemaître *et al.* [62] reviewed 19 studies in the literature which have reported the sensitivity performance of their methods. The average *Sen*=81.5% with average number of prtients is 27 and only five methods have achieved above these average values (in terms of *Sen* against the number of patients) and three of them are our proposed methods in Chapters 4, 5 and 6 (indicated in magenta, red and green, respectively). The other two methods were proposed by Puech *et al.* [346] and Peng *et al.* [45]. Once again, the graph indicates that there is no correlation between sensitivity and the number of patients. All our proposed methods achieved *Sen*>86% and although the method of Puech *et al.* [346] produced *Sen*=100%, it has poor *Spe*=43%.

Figure 7.4 shows the distribution of *Spe* against the number of patients from studies in the literature (only those studies have reported *Spe*). Similar to *Sen* Lemaître *et al.* [62] reviewed19 studies in the literature. The average *Spe* of the 19 methods is 82.2% which is slightly higher than the *Sen* on the same average number of patients. Based on this average value only four methods have achieved 'above the average zone' and three of them are our proposed methods in Chapters 4, 5 and 6 (indicated in magenta, red and green, respectively). The other method which produced the highest specificity within the 'above the average zone' was proposed by Peng *et al.* [45].

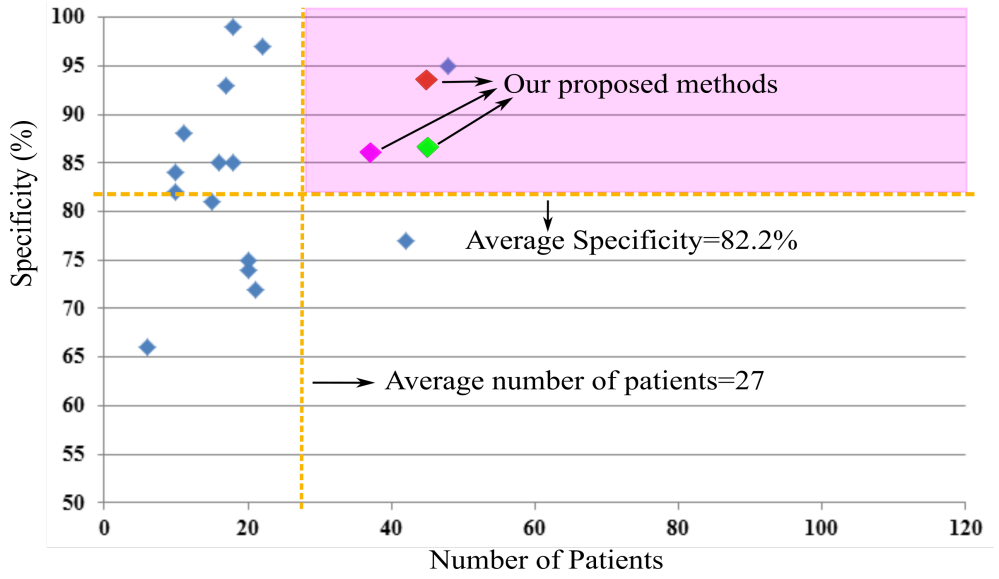

FIGURE 7.4: A scatter plot representation of the *Spe* against the number of patients.

TABLE 7.4: The average performance metric for multiparametric and monoparametric studies.

Metric(%)	**Multiparametric**	**Monoparametric**
$\boldsymbol{A_z}$	87.6	87.6
CA	84.5	85.3
Sen	76.6	85.9
Spe	79.7	84.5
# patients	37	36

7.4 The Average of the State-of-the-art

This section presents the average performances based on the different categories and streams of the CAD-PC studies in Table 7.1, Table 7.2 and Table 7.3. The first stream is based on the modality used and the second stream is based on the strength of a magnetic field used to acquire the dataset. In the first stream, studies are divided into multiparametric and monoparametric. The second stream categorises each study based on 1.5T or 3.0T. The main difference between these strengths is that 3.0T has better resolution allows a better quality for anatomical representation.

Table 7.4 shows the average of each performance metrics in mono and multiparametric studies. We found 26 CAD-PC multiparametric studies in the literature with an average number of patients is 37 per study. However, not all studies reported the

TABLE 7.5: The average performance metric for studies based on 1.5T and 3.0T datasets.

Metric(%)	**1.5T**	**3.0T**
$\boldsymbol{A_z}$	88.9	83.2
$\boldsymbol{CA}$	85.2	83.0
$\boldsymbol{Sen}$	84.4	73.2
$\boldsymbol{Spe}$	80.6	86.8
# patients	27	57

result of all performance metrics. The average A_z and CA per multiparametric study are 87.6% and 84.5%, respectively which are similar with monoparametric studies (87.6% and 85.3%, respectively). In contrast, the average sensitivity and specificity for each multiparametric study are 76.6% and 79.7%, respectively. In monoparametric studies,the average Sen=85.8% and Spe=84.5% which are higher than the average values in multiparametric studies.

On the other hand, Table 7.5 shows the average performance metrics in studies based on data acquired using 1.5T and 3.0T magnetic field. The results show that CAD-PC from datasets acquired using 1.5T produced higher results in all metrics except specificity. The average A_z, CA and Sen in these studies are 88.9%, 85.2% and 84.4%, respectively. Studies used higher spatial resolution datasets produced Spe = 86.8% which is 6.2% higher than those studies have used datasets with a lower spatial resolution. Studies based on 3.0T dataset produced lower Sen=73.2% and A_z=83.2% but similar CA=83%.

Based on the average values in table 7.4 and 7.5, it seems that CAD-PC based on monoparametric and 1.5T dataset produced better results in terms of A_z, CA and Sen. However, we would like to emphasise to the reader that this does not allow us to draw definitive conclusions because the average values were calculated based on different studies using different datasets. Nevertheless, this tells us about the trends of the current state-of-the-art. In a more robust comparison, several studies [46, 64, 196, 229] indicated that using multiparametric MRI leads to a better performance as more information about the tissues characteristics can be acquired. In those studies, they evaluate the performance results based on exactly the same dataset. Similarly, although studies based on 1.5T datasets produced better results on average, it does

not mean that using 1.5T datasets are better because practically an MRI scanner with 3.0T produce better spatial resolution allowing high quality vascular imaging.

7.5 Findings from Qualitative Comparisons

We would like make the following findings drawn based on the qualitative comparisons of the different studies in the literature:

1. It is difficult to make a quantitative comparison based on the results in Tables in 7.1, 7.2 and 7.3. Therefore, it is impossible to draw a definitive conclusion as to which method is superior because many factors (as previously discussed) come into play.

2. Regarding the optimal performance, although the method in [211] based on single modality produced the highest A_z=97%, several other studies [46, 64, 196, 229] showed that using multiparametric MRI data improved the performance results significantly compared to monoparametric MRI data.

3. In multiparametric studies combining features from T2-W and DCE produced the best A_z=97% qualitatively as reported in [21] whereas combining features from three modalities (T2-W + DCE + DWI) generated best A_z=95%.

4. For the similar reason in point number 1, it is difficult to draw a firm conclusion about which modality provides better performance for an individual method. Nevertheless, from the results in Tables in 7.1, 7.2 and 7.3, CAD-PC based on single modality (T2-W) produced better results qualitatively. However, Tiwari *et al.* [229, 350, 355] observed that MRS is more suitable modality to highlight cancer regions. In another study, Litjens *et al.* [46] observed that DWI modality was more suitable than DCE and T2-W to distinguish cancerous regions in their method.

5. In recent years, many multiparametric CAD-PC have been developed due to vast amounts of features which can be extracted from different modalities, and hence have attracted the attention of both radiologists and computer vision researchers.

7.6 Qualitative Comparison with Human Performance

Niaf *et al.* [231] reported preliminary results with radiologists performances for two and 15 years experience with A_z=80% and 86%, respectively. Litjens *et al.* [46] reported the range of 10 radiologist performances were between A_z=81% to 83%. From these studies, we have a general idea of a typical radiologist's performance in comparison to CAD-PC. In our case, the proposed method achieved A_z=93% which is significantly better (p <0.001) than human performance and some CAD systems based on multiparametric MRI [65, 231, 300, 353].

7.7 Summary

To summarise, this chapter highlights the following:

1. We have compared our results with the state-of-the-art qualitatively and our proposed methods achieved above the average performance in all metrics (A_z, CA, Sen and Spe).
2. Due to absence of public datasets, different modalities, different levels of evaluation and different studied zones of the prostate gland it is extremely difficult to make a quantitative comparison. Therefore this underlines one of the major ongoing deficiencies in the study of CAD-PC which need to be resolved.
3. We have briefly described some of the major CAD-PC studies in the literature.
4. From the compiled results shown in Table 7.1, 7.2 and 7.3, it is difficult to draw a conclusion which method is superior in the literature.

Chapter 8

Summary and Discussion

The major focus of this thesis is developing computer algorithms for prostate cancer detection in MRI. Many studies have considered the potential of CAD-PC and shown it to be quite promising. In this chapter, we will focus on the main areas of this thesis both from the point of view of the general audience and computer scientists. In addition, we will discuss the advantages, limitations, challenges and the prospects of CAD-PC in clinical practices.

8.1 Chapter 1: Introduction

In Chapter 1, we have presented the current data covering incidence of prostate cancer in the world, Europe and the UK. In general, although the survival rates have increased due to early detection the number of cases are expected to increase to 1.7 million in 2030 globally. However this figure is partly influenced by the increase in detection incidence from 'less developed countries'. This reflects the current screening methods (e.g. PSA and DRE) within these countries. The increase is also because more sophisticated screening methods will be available in the future which will enable clinicians to detect prostate cancer at a very early stage, which will improve the prognoses for those with the disease.

In terms of screening methods, PSA, DRE and TRUS are among the most common procedures available in most hospitals. Many studies have discussed the accuracy, sensitivity and specificity as well as several concerns such as over-diagnosis, over-treatment, invasiveness and post operative complications. It would be interesting

to investigate the variability among clinicians especially when performing DRE and TRUS methods as the accuracy for these methods is highly reliant on the clinician's experience (e.g. when taking sample tissues from the prostate gland).

In this chapter also we briefly discussed the role of MRI in prostate cancer. Currently, MRI is only used to further investigate the staging of prostate cancer already diagnosed and confirmed by the doctor. This means MRI assists clinicians by providing better prognosis and treatment planing for the patient. In contrast, MRIs have been used as a first reader setting in mammography and lung imaging. MRI technologies are expensive, and to our knowledge there is no study in the literature has investigated the performance of MRI as a first reader screening method. The use of multiparametric MRI is becoming popular both among the radiologists and computer scientists. Nevertheless, the lack of standardised protocol among institutions is currently a major issue. Therefore, currently MRIs are used to support the development of CAD-PC by providing a vast amount of data.

8.2 Chapter 2: Texture analysis

In Chapter 2, we briefly reviewed different aspects of texture analysis and its applications in different fields. In the last 30 years, many different types of texture analysis methods and techniques have been developed and successfully applied in research areas. However, due to the degree of complexity and diversity it is impossible to give a universal definition of textures [66]. Therefore, many studies in the literature define textures based on the application and problem domain. In addition, textures also can be defined from different perspectives such as statistical, psychological and image and signal processing.

According to Ojala *et al.* [74, 75], there are four main problem domains in texture analysis: (a) texture classification, (b) texture segmentation, (c) texture synthesis and (d) shape analysis. In texture classification, training is needed to build a predictive model before the model is used to classify the test data. In contrast, texture segmentation is a process of grouping or clustering pixels based on their properties such as contrast, homogeneity, etc. This means, pixels that represent the same texture will be grouped together. On the other hand, texture synthesis is basically a method to create textures artificially and a process of computationally constructing a large

digital image from a small digital sample image [91, 94]. As for shape analysis it is a process to capture the geometric structures and contours of an image.

In texture analysis techniques, we have briefly reviewed four main methods covering statistical, structural, signal processing and model based. Statistical methods exploit the information of spatial distribution and intensity of pixel values in an image such as the grey level co-occurrence matrix introduced by Haralick *et al.* [108]. Geometrical methods attempt to capture the structural properties of the textures according to some placement rules such as orientation and distance [116]. In signal processing methods, the characteristics of a texture is computed using a set of filters which are designed to stimulate visual perception and spatial frequency in the human brain. Each filter is uniquely designed to capture a particular texture property. Finally, model based methods try to model the appearance of the texture by considering the relationship of each pixel with its neighbouring pixels.

We also presented some of the most popular texture descriptors for texture representation in CAD-PC studies. In Section 2.4, the following features were clarified: (a) first-order statistical features, (b) second-order statistical features, (c) local binary pattern (d) filter banks, (e) textons (f) gradient based features, (g) Tamura's features and (h) fractal based features. Each of these features was explained in a step-by-step manner in Chapter 2. Since these features were first introduced some time ago, many authors have made different variations and it is impossible to review and include every variant in this thesis.

Although many texture descriptors for texture representation are described in the literature, to our knowledge no study that has used all these features and evaluated them individually as to which feature performed best for prostate cancer detection in MRI. Similarly, there is no CAD-PC study which has attempted to investigate how certain features performed on different CAD-PC methods. This would have given an indication as to which of these components (features or method framework) has stronger influence on the performance results. In many CAD-PC studies, features were selected because of their excellent performance as a general texture classification rather than for clinical reasons.

8.3 Chapter 3: Computer Aided Detection Systems

We briefly discussed the general framework of typical CAD-PC both mono and multiparametric in Chapter 3. Although there are many CAD-PCs have been developed in the literature, these methods shared some common workflows such as pre-processing, feature extraction, training and testing. In pre-processing step, most CADs employed noise filtering and normalisation to remove artifacts caused by the magnetic coil and to standardise pixel intensity, respectively. In some cases, depending on the data, some CADs involved bias correction to remove the hyper-intense signal around the coil of an image. Image registration techniques are popular among CADs based on multiparametric MRI. In feature extraction, image based features (both on region and pixel-based features) are among the most popular information used to train classifiers. Image based features include texture-based, edge-based, histogram-based, intensity-based, etc. Another set of features are those extracted from DCE modality called DCE-based features. In contrast to image-based features, these features represent the different metabolites levels in prostate tissues based on whole-spectra approach, semi-quantitative approach and quantitative approach. Using the same approaches, MRSI-based features can be extracted from MRSI modality. However, MRSI-based features not only represent metabolites levels but metabolic activity occurring in the malignant and benign regions within the prostate gland.

From dimensionality reduction point of view, principal component analysis and statistical tests are the first choice for most researchers in CAD-PC studies. On the other hand, in training and testing phases, the SVM classifier using RBF kernel is the most frequent machine learning algorithm used to distinguish benign and malignant instances. However in the last five years the RF classifier has gained popularity tremendously among the medical image analysis community. Many studies have shown the effectiveness of SVM classifier after parameter optimisation. In contrast, most CAD-PC studies have used the RF classifier did not perform parameter optimisation as the method itself is already robust [46]. However, it might be interesting to make a quantitative comparison between these classifiers (since these are the most popular classifiers in CAD-PC studies) before and after parameter optimisation. In this thesis we have found that SVM did not perform well mainly due to default parameters in WEKA. On the other hand, the RF classifiers performed well both in

Chapters 5 and 6. In addition, we discovered two more classifiers which performed well in both methods (Chapters 5 and 6) are the MLP and BNet classifiers.

8.4 Chapter 4: Detection and Localisation of Prostate Cancer using Intersection of Binary Segmentations

In Chapter 4, we proposed a new unsupervised CAD method for prostate cancer detection using the intersection of binary segmentations. For this purpose, we extracted four different image features based on Gaussian and median filtering techniques. The first two image features were computed using the conventional Gaussian (G_1) and median (M_1) filtering. The third image feature (F_1) was computed by calculating the image magnitude between G_1 and M_1 before applying a robust smoothing method in [273]. On the other hand, the fourth image feature (F_2) was calculated by computing the probability image from each G_1 and M_1. Subsequently, the image magnitude (F_2) is computed from the resulting probability images. In the image segmentation phase, each of the extracted image features are segmented individually using the modified FCM algorithm of Chen and Zwiggelaar [256] which resulted in four separate binary segmentations. Using these binary images, we took the overlapping segmented regions as the most cancerous areas. Finally to remove false positives we performed erosion on the resulting image. Evaluation results showed that the proposed method achieved similar performance to the other CAD-PC methods in the literature. The main advantage of this method is the fact that it uses only four feature images but still can produce similar results compared with the other methods which use a large number of features. Nevertheless, we would like to emphasise to the reader that the numerical results are based on different datasets and evaluation levels.

Many of the proposed CAD methods in the literature used machine learning techniques. In comparison, our method bypasses the training phase in a typical CAD framework. Technically, supervised CAD-PC methods could be time consuming depending on the classifiers. For instance, the SVM and MLP classifiers are slower in comparison to the NB and BNet classifiers. In fact, using classifiers such as SVM and k-NN requires the parameter optimisation step which is quite time consuming (depending on the number of training and testing data). In terms of complexity, supervised methods tend to be more complex as they require feature selection or

dimensionality reduction particularly those methods involving a large number of features. In addition, the training phase adds another layer of complexity especially when using a complex classifier such as the MLP. In terms of performance most supervised CAD methods are theoretically more robust as the method itself has the opportunity to learn and build a model of the training data. However it is hard to draw such conclusion due to limited number of unsupervised CAD-PC studies. Although Liu *et al.* [234] reported 89%, 88% and 90% accuracy, sensitivity and specificity, respectively, there is no qualitative comparison made by the authors against the other supervised methods in the literature.

8.5 Chapter 5: Multifeature and Unsupervised Machine Learning Algorithms in Prostate Cancer Detection

In contrast to the proposed method in the previous chapter, in Chapter 5 we investigated 215 texture descriptors covering different window sizes, filters, techniques and orientations. Furthermore, we employed 11 machine learning techniques ranging from tree-based, Bayesian-based, function-based to network-based classifier. Some of the classifiers such as the SVM and RF are the most popular machine learning algorithm in CAD-PC studies. The proposed method consists of five phases: (a) noise filtering (b) feature extraction (c) feature scaling (d) feature selection (e) data classification (training and testing). In the noise filtering phase, we combined median and anisotropic diffusion filtering methods to remove noise as well as to retain textures. Subsequently, a set of 215 features were extracted comprising first- and second-order statistics, filter bank, Tamura's features and gradient based features. In the next phase, each feature vector was normalised to [0-1] followed by feature selection to eliminate unnecessary features and reduce data dimensionality. In the final phase, 11 classifiers were trained and each predictive model was used in the testing phase.

Evaluation results showed that the proposed method achieved similar results to the current state-of-the-art in terms of A_z, CA, Sen and Spe. In this chapter we also investigated the effect of feature selection on the classification performance. Results showed a significant improvement for most classifiers when employing feature selection. However most tree-based and the SVM classifiers produced less than 5% improvement.

Furthermore, we investigated the effect of window sizes on feature's performance individually. Our evaluations results showed that the second-order statistical features are more effective at smaller resolutions whereas the filter bank performed well from 3×3 until 13×13 before their performances decreased at much larger window sizes. Gradient based features are among the most consistent texture descriptors which performed well in all tested window sizes (from 3×3 to 19×19). For first-order statistical and Tamura's features, only certain features are discriminative (e.g. percentile 75% and Tamura's contrast). To our knowledge our study is the first to investigate the effects of window sizes on the feature's performance individually. Nevertheless, the evaluation was only based on CfsSubsetEval [327] attribute evaluator and the GreedyStepwise search method in WEKA [2] which is one of the limitations of this study. Another limitation is that we were unable to compare our results quantitatively with the other methods in the literature mainly due to absence of public datasets. Since our study employed 11 different classifiers to get the best possible results, performing parameter optimisation for each of the classifiers is time consuming and computationally expensive (therefore all parameters in this study were left on the default settings in WEKA). However, we have made an initial investigation regarding to this issue on the ADTree, k-NN and RF classifiers and included our preliminary results in Section 5.6 and 6.7. Finally, the proposed method showed that it is possible to develop a monoparametric CAD method for prostate cancer detection with similar performance to the ones based on multiparametric MRI.

8.6 Chapter 6: A Texton Based Approach in Prostate Cancer Detection

Our final CAD method proposed in Chapter 6 is based on textons. Similar to textures, many studies argued that it is hard to define a universal definition of textons due to lack of a good mathematical model [170]. However, textons can be seen as a representative of micro-structures in natural images and are considered as the atoms of pre-attentive human visual perception [169]. In our study, we defined textons as the k-means' cluster centers. In our proposed method, there are two main advantages of using a texton based approach: firstly it bypasses the conventional feature extraction methods based on filtering and secondly it avoids the typical dimensionality reduction

methods such as feature selection which both can be time consuming. To our knowledge this is the first texton based CAD approach in the literature applied in relation to prostate cancer detection.

The proposed method consists of the following steps: (a) pre-processing, (b) construct texton dictionary, (c) feature extraction, (d) training and testing. We used the same median and anisotropic diffusion filtering techniques in Chapter 5 in the pre-processing phase. To construct the texton dictionary we did not use filter bank as originally proposed by Varma and Zisserman [1]. Instead we followed the the later study by them in [161] which clusters benign and malignant patches directly from the original image pixels. In the feature extraction phase, each pixel is represented as a histogram treated as a feature vector. The constructed histogram for each pixel consists of the frequency of the neighbouring texton occurrence within the ws (or patch size) including the texton at the central pixel. Subsequently we employed 8 classifiers to build predictive models and test the models on the unseen cases.

Evaluation results show the proposed method achieved similar results with the state-of-the-art in all performance metrics. Indeed, most texton based approaches are rely on two parameters ws and txt. In our experiments we found $ws = 9 \times 9$ and 11×11 with 6 and 16 textons produced the best results for most classifiers. On the other hand, the BNet, RF and C4.5 classifiers are the best three machine learning algorithms produced A_z>87%. This chapter further supports our findings in Chapter 5: firstly in terms of using suitable window sizes, secondly the proposed method further suggests that it is possible to develop CAD methods based on T2-W MRI which can achieve similar performance to the ones based on multiparametric MRI and thirdly, quantitative comparison in both studies (Chapters 5 and 6) showed that the BNet produced the best A_z value as an individual classifier.

Finally, one of the limitations of our study is we did not perform parameter optimisation of the machine learning algorithms employed in Chapters 5 and 6 (except the $k-$NN, ADTree and RF classifiers) but used the default parameter settings in WEKA [2]. However, despite using default parameter settings, our proposed methods showed that A_z>90% which is similar with most methods in the literature. The results of varying the parameters for the RF and ADTree classifiers shown that similar results were achieved for both performance metrics (A_z and CA). However, at a very small values of nB (e.g. $nB = 1$) and rF ($rF = 5$), at least 3% to 5% difference in comparison to the results produced at a given parameter value in WEKA. Our preliminary results of parameter optimisation for the ADTree, k-NN and RF classifiers indicated

that the default parameter values given in WEKA are good enough as initial settings investigate the classifier's performance.

8.7 Chapter 7: Qualitative comparisons

Chapter 7 presented qualitative comparisons between our proposed methods and the state-of-the-art covering four different performance metrics A_z, CA, Sen and Spe. We presented 38 CAD-PC studies in the literature including our methods and rank them based on A_z, CA, Sen and Spe accordingly. The highest A_z=97%, CA=89%, Sen=100% and Spe=99% produced by the methods by Vos *et al.* [21], Liu *et al.* [234], Puech *et al.* [346] and Matulewicz *et al.* [356], respectively. Although these numbers look high the averages are A_z=87.5%, CA=84.9%, Sen=81.5% and Spe=82.2% based on 38 studies including our methods. Qualitative comparisons showed that our proposed methods produced similar results in all four performance metrics.

If considering the average A_z against the average number of patients only six methods [45, 46, 211, 348] in the literature managed to achieved the 'above the average zone' including our proposed methods in Chapters 5 and 6. However, in terms of average CA versus average number of patients only our methods in Chapters 4 and 6 produced results within that zone. On the other hand, two methods [194, 346] and all of our proposed methods produced sensitivity results which are 'above the average zone'. Similarly, all our proposed methods produced above the average in terms of specificity versus the number of patients together with the method of Peng *et al.* [45].

In general, the CAD-PC studies can be seen in two different streams: firstly CAD methods based on mono and multiparametric and secondly whether data is acquired using scanner 1.5T or 3.0T magnetic field strength. In Chapter 7, we calculated the average performances between these categories and found out that there is no significant improvement between mono and multiparametric in terms of A_z and CA. However, monoparametric MRI produced better Sen=85.9% and Spe=84.5% in comparison to multiparametric Sen=76.6% and Spe=79.7% with 37 and 36 average number of patients, respectively. In terms of datasets, methods based on 1.5T datasets produced better average results in all metrics except specificity where methods based on 3.0T produced average Spe=86.8%. Nevertheless, these comparisons should be carefully interpreted because each method was evaluated based on different datasets.

As discussed in the previous chapter, quantitative comparison is difficult due to differences in the type of datasets, absence of public datasets, studies being conducted at different regions and evaluation at different levels. These issues remain problematic making the study of CAD-PC complex in terms of building a solid foundation for the state-of-the-art as well as qualitative comparison. Furthermore, the limited number of quantitative comparisons between CAD-PC against radiologists performance makes it hard to draw a conclusion as to the extent to which the computer systems can help radiologists in diagnosis decision making.

Chapter 9

Conclusions and Future work

The conclusions of the work covered in this thesis are given in this chapter followed by proposals for future work which would build on the the findings this thesis. This covers how the aim and objectives of this thesis were achieved. Finally, the novel contributions of this thesis are highlighted and a list of publications is presented in the final section of this chapter.

9.1 Conclusions

In this thesis, we have presented three CAD methods for prostate cancer detection within the PZ based on a single modality T2-W MRI using image processing and machine learning techniques. Evaluation results showed that all proposed methods achieved similar results compared with the-state-of-the-art in terms of A_z, CA, Sen and Spe in both CAD-PC based on mono and multiparametric MRI. Moreover, the proposed methods outperformed some of the methods using 3.0T datasets which have better imaging resolution in comparison to 1.5T datasets. Therefore this answers our research question raised in Chapter 1 as to whether it is possible to develop a computer algorithm based on single modality and achieve similar performance to those based on multimodality. We are aware that T2-W MRI alone is insufficient in developing a more robust CAD-PC system as multimodality MRI can provide more informative data (e.g. physiological tissue characteristics and metabolites composition) which are not available in a conventional MRI. Nevertheless, this study identifies a set of discriminant texture descriptors which can be combined with features from the other modalities, hence provide a solid basis for a CAD-PC based on multimodality.

We started by investigating the epidemiology of prostate cancer including prostate cancer incidence and mortality, the current screening methods for prostate cancer detection including their advantages and disadvantages, the prospect of mono and multiparametric MRI in prostate imaging and how CAD in MRI can be beneficial to clinicians and patients. Following this investigation, we carried out a literature review on MRI image analysis and studied existing approaches in the field which established a foundation for our work in this thesis. It was clear that the development of CAD-PC is an important and active field among the research community in medical image analysis. In comparison to other the CAD methods for breast, brain, bone, and lung imaging, the number of CAD-PC is relatively small. Moreover, many CAD-PC studies were tested based on small datasets. Therefore, further developments are needed to investigate the prospects of CAD as a first reader as a diagnostic tool in conjunction with radiologists in hospitals. Based on these, we concluded that there is an urgent need to develop CAD methods for prostate cancer detection whether based on mono or multiparametric MRI both for detection and staging.

Subsequently (Chapter 2), we investigated the most popular texture descriptors used in CAD-PC which are first- and second-order statistical features. Other features such as LBP, fractal, gradient and intensity based are also among the most popular features used in CAD-PC studies. We also investigated the general framework of CAD-PC based on the existing studies in the literature (Chapter 3). Based on the 35 CAD-PC studies reviewed in [62], we concluded that a typical CAD-PC consists of three main phases: (a) pre-processing, (b) feature extraction and (c) training and testing classifiers. However, in some cases additional phases might be needed such as post-processing and image regularisation or registration. This gave the basis of framework for the proposed methods in this thesis. In Chapter 2, we made an initial step to achieve the project's second and fifth objectives by investigating different texture descriptors used in the literature and Chapter 3 gave us an idea of a typical design of a CAD-PC arcitecture.

In our first method in Chapter 4, we have successfully developed an unsupervised CAD method using four texture descriptors and used the modified FCM algorithm to cluster each image feature individually. The resulting binary segmentations were then stacked and the overlapping regions are taken as malignant regions. In this method, we found that the use of probability image was effective to quantify the probability of each pixel being malignant or benign. On the other hand, image magnitude was used to capture consistent behavior of the texture. For noise removal, we used median

and Gaussian filters and erosion technique was employed to remove false positives and negatives by $\approx 20\%$. As a result, the proposed method achieved 86%, 83% and 96% accuracy, sensitivity and specificity, respectively. In this chapter, we partly achieved our fifth objective by developing an unsupervised CAD-PC. In addition, we have investigated the texture appearance of malignant and benign regions within the PZ using features smoothed by median and Gaussian filters (hence we partly achieved our second objective). The first objective of this thesis was also achieved in this chapter as we have developed a 2D model to automatically estimate the boundary of the PZ.

Further, we developed a supervised CAD method in Chapter 5. For this purpose a sophisticated noise removal of anisotropic diffusion filtering method in conjunction with median filtering were used to remove unwanted artifacts from our dataset. Since there are many different texture descriptors which have been developed in recent decades, it is time consuming to implement each of them. Nevertheless, we investigated 215 texture descriptors covering different resolutions, filters, techniques and orientations. For classification purposes, we employed 11 machine learning algorithms to achieve the best possible results. In this method, we addressed five different questions: (a) does ws affect the CAD performance both in terms of A_z and CA?, (b) does the selection of ws affect the performance of the texture descriptor itself?, (c) can we produce a general guideline for using ws in CAD-PC studies?, (d) which features performed well in CAD-PC? and (e) which machine learning algorithms performed well and consistent in CAD-PC studies?. Our experimental results suggested the following answers: (a) the selection of ws has a significant effect on CAD-PC methods in terms of A_z and CA, (b) the selection of ws has a significant effect on the performance for most texture descriptors, (c) our experimental results suggested that 9×9 and 11×11 generally have significantly better results in comparison to the other resolutions both in terms A_z and CA (d) gradient based features performed well in most ws whereas filter bank, Tamura's contrast feature and percentile 75% are also highly recommended at medium size resolution, and (e) combining different classifiers to decide a class boundary often produced better results. However, individually the BNet, ADtree, RF, and MLP produced better results in comparison to the other classifiers tested in our experiments. As a result, the proposed method in Chapter 5 produced A_z=92.7% when several classifiers were used to draw a class boundary and individually the BNet classifier produced the best A_z=90%. This chapter extended the achievement of the project's fifth objective via the development of supervised CAD-PC. Furthermore, we also extended the achievement of our second objective as 215 texture descriptors

were investigated. Based on our experimental results, we identified the most discriminant features which can differentiate benign and malignant regions. Furthermore, we found that $ws = 9 \times 9$ and 11×11 produced the best A_z and CA for most classifiers. Therefore, we achieved our third and fourth objectives in Chapter 5.

In our third method (Chapter 6) we investigated how textons can be used as robust texture descriptors in prostate cancer detection. For this purpose we developed a CAD method based on the occurrence of textons within the neighbouring pixels. In this method we take the definition of textons as the resulting centroids from the k-means clusering algorithms. The method contains three main phases: constructing texton dictionary, feature extraction and training and testing. To our knowledge this is the the first CAD-PC based on textons. Our study (in Chapter 6) highlighted two main advantages of using textons: firstly, it bypasses the typical feature extraction techniques (e.g. filtering and convolution) and secondly dimensionality reduction (e.g. feature extraction) is not needed. On the other hand, visual comparison showed that extracted textons from benign regions are smoother in comparison to the ones extracted from malignant regions. Evaluation results showed the BNet classifier (A_z=92.8%) once again outperformed the other classifiers, individually followed by the RF classifier which achieved A_z=89.5%. Our experimental results suggest that $ws = 9 \times 9$ and 11×11 with $txt = 6$ and 16 produced the best results in terms of A_z and CA. This further confirms our experiments results in Chapter 5 in terms of the selection of ws in CAD-PC studies. Moreover, experimental results in Chapter 5 and 6 suggest that there is no significant difference between the results achieved at 9×9 and 11×11 in terms of A_z and CA. Chapter 6 completed the second,fourth and fifth objectives of this project because: (a) a texton based approach was used to further investigate the texture appearance of malignant and benign regions (visually textons extracted from malignant regions look rougher), (b) experimental results suggest that best results were achieved at 9×9 amd 11×11 which further support our initial findings in Chapter 5 and (c) another supervised CAD-PC was developed using textons. The final objective of this project was partly achieved in Chapters 4, 5 and 6 as proposed methods were evaluated based on 275 and 418 MR images of 37 and 45 patients, respectively.

In terms of qualitative comparison with the state-of-the-art, the results in Chapter 7 suggested that our methods have similar capability in all performance metrics. The comparisons were made against 35 CAD studies in the literature covering seven different aspects which are A_z, CA, Sen, Spe, number of patients, modality (e.g. mono

or multiparametric MRI) and types of datasets (e.g. 1.5T or 3.0T). At present, valid quantitative comparisons of CAD-PC studies are impossible due to the differences in datasets such as the number of patients (or dataset size) and different modalities. In fact, the absence of publicly available datasets makes it extremely difficult to compare the existing methods. As a result, for most existing CAD-PC described in the literature can only make qualitative comparisons. On the other hand, although several studies in [46, 64, 196, 229] claimed multiparametric MRI produced significantly better results than monoparametric, our analysis comparison in Chapter 7 showed that there is no significant difference in terms of A_z and CA on average, whereas in terms of Sen and Spe CAD-PCs based on single modality tends to produce better results. However, these are the average results based on different studies, which means results were produced based on different datasets. The final objective of this thesis was fully completed as extensive qualitative comparisons were made in Chapter 7 against 35 CAD-PC studies mainly covering different performance metrics and datasets.

From a clinical point of view this thesis highlighted the following advantages: (a) a fully automatic CAD-PC can minimise the workload of radiologists in analysing MRI images, (b) the development of CAD-PC not only can help the detection process faster, but robust and consistent, (c) CAD-PC can provide a visual representation to clinicians as to which regions of the prostate have high probability of being malignant, helping more accurate biopsy sampling (d) CAD-PC can provide a second opinion to the radiologist, hence improving detection accuracy. On the other hand, from a computer vision point of view this thesis highlighted the following advantages: (a) we have highlighted the most discriminant texture descriptors from the 215 features extracted, (b) this thesis also provides general guidelines regarding window size when extracting features including how resolution size affects the performance both of the feature and the method itself and (c) our experiment results in Chapter 5 and 6 suggested that the BNet and RF classifiers provide the best results in terms of A_z and CA.

9.2 Future Work

Based on our investigation conducted in this thesis and analysis of our experimental results, we suggest the following future work:

1. It should be noted that all the CAD-PC developed in this thesis were only tested within the prostate's PZ region. Therefore the main recommendation for future work of this thesis is to develop or modify the methods and evaluate them covering the whole prostate gland. This includes investigating benign and malignant regions within the CZ by finding the most discriminant texture descriptors.

2. Since all machine learning algorithms employed in this study were left on the default parameter settings in WEKA [2], future work may investigate the effects of parameter optimisation. However, preliminary results of the k-NN, RF and ADTree classifiers showed that the default parameters given in WEKA are fairly good as an initial guideline on selecting a classifier in CAD-PC.

3. Although we have investigated more than 200 texture descriptors, there are still many features to be investigated such as histogram based features, Scale-invariant feature transform (or SIFT), etc.

4. Since this thesis only focuses on a single modality T2-W MRI, the future work of CAD-PC is to investigate the performance of the proposed methods if features from other modalities such as DCE and DWI are combined.

5. A study of quantitative comparison can be seen as one of the main challenges in future direction of CAD-PC studies.

9.3 Novel Contributions

The main contributions and novel aspects of this thesis are summarised as follows:

1. This thesis presents a substantial literature review regarding CAD-PC studies covering the statistics incidence and mortality, current status of prostate cancer from clinical point of view, texture analysis in mono and multiparametric MRI, more than 200 texture descriptors, a general framework of CAD-PC, 11 machine learning algorithms and 35 CAD-PC studies.

2. A new unsupervised CAD-PC approach which is based on combining multiple binary segmentation using only four simple image features. The novelty of this method lies in the approach which combines simple features (this is similar to a forest of weak classifiers which together provide strong results), for the first

time and applied to prostate T2-W using one modality. To our knowledge, no existing methods in the literature have used the technique of finding cancer regions by taking the overlapping binary segmentation extracted from a small number of image features.

3. A new supervised CAD-PC using 215 texture features covering different window sizes, filters, techniques and orientations from T2-W MRI only. This means, this thesis investigated more feature options and some of them have not yet been applied in CAD-PC. To our knowledge the previously used largest number of 2D texture descriptors in the literature using only T2-W MRI was in a study conducted by Viswanath *et al.* [295] (110 texture descriptors) and 83 texture descriptors by Tiwari *et al.* [65]. Using multiparametric MRI Niaf *et al.* [231] extracted 140 texture descriptors from T2-W MRI, diffusion-weighted imaging (DWI), and dynamic contrast enhanced (DCE).

4. We quantitatively compared 9 classifiers' performances with two additional combined classifiers (11 classifiers in total). Again, to our knowledge the largest number of classifiers used in the literature in CAD-PC is the study conducted by Niaf *et al.* [231] (4 classifiers), and Litjens *et al.* [46] and Ozer *et al.* [198] employed 3 classifiers. The classifiers' performances were further tested in our experiment in Chapter 6 and some classifiers produced similar results. This provides a general guideline for future researchers with suggested options referred for the selection of classifiers.

5. Since this study involved a large number of texture descriptors, we evaluated all features individually and combined them to improve performances of the proposed method. By evaluating them, we are able to determine which features individually give the best performances on each classifier. The analysis of our experimental results generated a list of discriminative features for every window size which can be beneficial to the research community and future researchers regarding the selection of texture descriptors in CAD-PC studies.

6. We investigated the effect of different window sizes on the performance of the proposed method as well as the performances of the features individually. To our knowledge, none of the current CAD-PC in the literature have reported quantitatively the effect of window sizes on performance. Many studies [65, 231, 233, 263, 296] selected window size without a qualitative justification. In this thesis we investigated 9 different window sizes (from 3×3 to 19×19) and

qualitative comparison showed that $ws = 9 \times 9$ and 11×11 produced the best results. This also provides a general guideline for future researchers regarding the likely best options for window size in CAD-PC studies.

7. A new supervised texton based CAD-PC approach. This is the first CAD method which has investigated the use of textons in classifying benign and malignant tissues within the prostate. To our knowledge there is no CAD which has investigated textons in prostate cancer detection particularly in MRI.

8. The proposed texton based CAD-PC approach learns directly from image pixels without the need to use a filter bank. In comparison, most CAD-PC in the literature computed large number of texture descriptors via filtering, and this is computationally expensive. In fact, computing a large number of texture descriptors can also lead to an additional step such as feature selection or dimensionality reduction.

9. An extensive qualitative comparison with the state-of-the-art covered seven different aspects: A_z, CA, Sen, Spe, number of patients, modality (e.g. mono or multiparametric MRI) and types of datasets (e.g. 1.5T or 3.0T). This provides a summary for the research community regarding the current trends in CAD-PC studies.

9.4 List of Publications

A list of publications resulting from this thesis is shown as follows:

Journals

1. **A. Rampun**, Z. Chen, P. Malcolm, B. Tiddeman, and R. Zwiggelaar. Computer-aided diagnosis: detection and localization of prostate cancer within the peripheral zone. *International Journal for Numerical Methods in Biomedical Engineering*, in press (2015). ISSN 2040-7947. doi: 10.1002/cnm.2745. URL:http://dx.doi.org/10.1002/cnm.2745. (Accepted: IF=2.052)

2. **A. Rampun**, P. Malcolm and R. Zwiggelaar. Detection of Prostate Abnormality within the Peripheral Zone Using Grey Level Distribution. *Journal of Image and Graphics*, 2(1), pp. 15–21, 2014.

3. **A. Rampun**, L. Zheng, P. Malcolm, B. Tiddeman and R. Zwiggelaar. Computer Aided Detection of Prostate Cancer in T2-Weighted MRI within the Peripheral Zone. *Physics in Medicine and Biology*, 2016 (resubmitted).

4. **A. Rampun**, L. Zheng, P. Malcolm, B. Tiddeman and R. Zwiggelaar. Computer Aided Diagnosis of Prostate Cancer: A Texton Based Approach. *Medical Physics*, 2016 (under review)

Conference papers

1. **A. Rampun**, H. Strange and R. Zwiggelaar. Texture segmentation using different orientations of GLCM features. In *proceedings of the 6th International Conference on Computer Vision/Computer Graphics Collaboration Techniques and Applications*, MIRAGE '13, Berlin, Germany, pp. 519–528, June 2013.

2. **A. Rampun**, Z. Chen, P. Malcolm and R. Zwiggelaar. Detection and localisation of prostate abnormalities. In *proceedings of the 3rd International Conference on Computational Mathematical Biomedical Engineering*, CMBE'13, Hong Kong, pp. 205–208, December 2013.

3. **A. Rampun**, P. Malcolm and R. Zwiggelaar. Detection of Prostate Abnormality within the Peripheral Zone using Local Peak Information. In *proceedings of the 3rd International Conference on Pattern Recognition Applications and Methods*, ICPRAM'14, Anger, France, pp. 510–519, April 2014.

4. **A. Rampun**, P. Malcolm and R. Zwiggelaar. Detection and Localisation of Prostate Cancer within the Peripheral Zone using Scoring Algorithm. In *proceedings of the 16th Irish Machine Vision and Image Processing*, IMVIP'14, Londonderry, Northern Ireland, pp. 75–80, August 2014.

5. **A. Rampun**, P. Malcolm and R. Zwiggelaar. A Block-based Approach for Malignancy Detection within the Prostate Peripheral Zone in T2-weighted MRI. In *proceedings of the 2nd Bioimaging Conference*, BioImaging'15, Lisbon, Portugal, pp. 56–63, January 2015.

6. **A. Rampun**, P. Malcolm and R. Zwiggelaar. Computer-Aided Diagnosis Method for MRI-guided Prostate Biopsy within the Peripheral Zone using Grey Level Histograms. In *proceedings of the 7th International Conference on Machine Vision*, ICMV'15,Milan, Italy, pp. 56–63, January 2015.

7. **A. Rampun**, L. Zheng, P. Malcolm and R. Zwiggelaar. Classifying Benign and Malignant Tissues within the Prostate Peripheral Zone using Textons. In *proceedings of the 19th Medical Image Understanding and Analysis Conference*, MIUA'15, Lincoln, UK, pp. 138–143, July 2015.

8. **A. Rampun**, L. Zheng, P. Malcolm and R. Zwiggelaar. Computer Aided Diagnosis of Prostate Cancer within the Peripheral Zone in T2-Weighted MRI. In *proceedings of the 19th Medical Image Understanding and Analysis Conference*, MIUA'15, Lincoln, UK, pp. 207–212, July 2015.

Bibliography

[1] M. Varma and A. Zisserman. A statistical approach to texture classification from single images. *International Journal of Computer Vision*, 62(1):61–81, 2005.

[2] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. The weka data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1):10–18, 2009.

[3] Prostate Cancer UK. What is my risk? http://prostatecanceruk.org/prostate-information/what-is-my-risk, 2015. Accessed 16-July-2015.

[4] Cancer Research UK. Prostate cancer risks and causes. http://www.cancerresearchuk.org/about-cancer/type/prostate-cancer/about/prostate-cancer-risks-and-causes, 2015. Accessed 16-July-2015.

[5] R. Siegel, J. Ma, Z. Zou, and A. Jemal. Cancer statistics, 2014. *CA: A Cancer Journal for Clinicians*, 64(1):9–29, 2014.

[6] Prostate Cancer UK. Prostate cancer facts and figures. http://prostatecanceruk.org/information/prostate-cancer-facts-and-figures/, 2015. Accessed 16-April-2015.

[7] N. Howlader, A. M. Noone, M. Krapcho, J. Garshell, N. Neyman, S. Altekruse, C. Kosary, M. Yu, J. Ruhl, Z. Tatalovich, A. Cho, H. Mariotto, D. Lewis, H. Chen, E. Feuer, and K. Cronin. Seer cancer statistics review,1975-2010, national cancer institute. http://seer.cancer.gov/archive/csr/1975_2010/, 2013. Accessed 15-April-2015.

[8] R. Chou, J. M. Croswell, T. Dana, C. Bougatsos, I. Blazina, R. Fu, K. Gleitsmann, H. C. Koenig, C. Lam, A. Maltz, J. B. Rugge, and K. Lin. A review of the evidence for the u.s. preventive services

task force. http://www.uspreventiveservicestaskforce.org/uspstf12/prostate/prostateart.htm, 2010. Accessed 15-April-2015.

[9] Prostate Cancer UK. Key signs and symptoms of cancer. http://www.cancerresearchuk.org/cancer-info/spotcancerearly/, 2015. Accessed 13-June-2015.

[10] International agency for research on cancer. Globocan 2012: Estimated cancer incidence, mortality and prevalence worldwide in 2012. http://globocan.iarc.fr/Pages/fact_sheets_cancer.aspx, 2012. Accessed 2-August-2015.

[11] Cancer Research UK. Prostate cancer incidence statistics. http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/prostate-cancer/incidence#heading-Zero, 2015. Accessed 17-June-2015.

[12] S. B. Edge, D. R. Byrd, C. Compton, A. G. Fritz, F. L. Greene, and A. Trotti. *AJCC Cancer Staging Manual*. Springer, Chicago, 7th edition, 2010.

[13] A. M. De Marzo, E. A. Platz, S. Sutcliffe, J. Xu, H. Gronberg, C. G. Drake, Y. Nakai, W. B. Isaacs, and W. G. Nelson. Inflammation in prostate carcinogenesis. *Nature Reviews Cancer*, 7:256–269, 2007.

[14] Christine C. H. Lee, O. Akin-Olugbade, and A. Kirschenbaum. Overview of prostate anatomy, histology, and pathology. *Endocrinology and metabolism clinics of North America*, 40(3):565–575, 2011.

[15] F. H. Martini, M. J. Timmons, and R. B. Tallitsch. *Human Anatomy*. Pearson Benjamin Cummings, San Francisco, 7th edition, 2012.

[16] E. J. Halpern, D. L. Cochlin, and B. B. Goldberg. *Imaging of the prostate*. Martin Dunitz Ltd, London, 1st edition, 2002.

[17] G. Murphy, M. Haider, S. Ghai, and B. Sreeharsha. The expanding role of MRI in prostate cancer. *American Journal of Roentgenology*, 201(6):1229–1238, 2013.

[18] W. K. Oh, M. Hurwitz, A. V. D'Amico, J. P. Richie, and P. W. Kantoff. *Biology of Prostate Cancer*. BC Decker, Harvard, 6th edition, 2003.

[19] Prostate cancer UK. Understanding the psa test. a guide for men concerned about prostate cancer. http://prostatecanceruk.org/media/41628/2782-understanding-psa-test.pdf, 2014. Accessed 4-August-2015.

[20] F. H. Schroder, J. Hugosson, M. J . Roobol, T. L Tammela, S. Ciatto, V. Nelen, S. Ciatto, V. Nelen, M. Kwiatkowski, M. Lujan, H. Lilja, M. Zappa, L. J. Denis, F. Recker, A. Berenguer, L. Maattanen, C. H. Bangma, G. Aus, A. Villers, X. Rebillard, T. van der Kwast, B. G. Blijenberg, S. M. Moss, H. J. de Koning, and A. Auvinen. For the erspc investigators. screening and prostate-cancer mortality in a randomized european study. *New England Journal of Medicine*, 360:1320–1328, 2009.

[21] P. C. Vos, T. Hambrock, J. Barentsz, and H. Huisman. Computer-assisted analysis of peripheral zone prostate lesions using T2-weighted and dynamic contrast enhanced T1-weighted MR. *Physics in Medicine and Biology*, 55(6):1719–1734, 2010.

[22] Ontario ministry of health and long term care. Prostate cancer and the psa test. http://www.health.gov.on.ca/en/public/programs/cancer/prostate/, 2012. Accessed 4-August-2015.

[23] S. F. Shariat and C. G. Roehrborn. Using biopsy to detect prostate cancer. *Reviews in Urology*, 10(4):262–280, 2008.

[24] American Society of Clinical Oncology (ASCO). Digital rectal exam (dre). http://www.cancer.net/navigating-cancer-care/diagnosing-cancer/tests-and-procedures/digital-rectal-exam-dre, 2014. Accessed 24-July-2015.

[25] R. Palmerola, P. Smith, V. Elliot, C. T. Reese, F. B. Mahon, L. E. Harpster, N. Icitovic, and R. D. Raman. The digital rectal examination (dre) remains important - outcomes from a contemporary cohort of men undergoing an initial 12-18 core prostate needle biopsy. *Canadian Journal of Urology*, 19(6):6542–6547, 2008.

[26] W. J. Catalona, J. P. Richie, F. R. Ahmann, M. A. Hudson, P. T. cardino, R. C. Flanigan, J. B. deKernion, T. L. Ratliff, L. R. Kavoussi, and B. LDalkin. Comparison of digital rectal examination and serum prostate specific antigen in the early detection of prostate cancer: results of a multicenter clinical trial of 6,630 men. *Journal of Urology*, 151(5):1283–1290, 1994.

[27] Prostate Cancer UK. How prostate cancer is diagnosed. http://www.nhs.uk/ipgmedia/National/Prostate%20Cancer%20UK/assets/Howprostatecancerisdiagnosed(PCC).pdf, 2014. Accessed 27-July-2015.

[28] Bupa Health Information Team. Transrectal ultrasound-guided prostate biopsy. http://www.bupa.co.uk/individuals/health.../t/transrectal-prostate-biopsy, 2013. Accessed 13-July-2015.

[29] D. Svetec, K. McCabe, S. Peretsman, E. Klein, H. Levin, S. Optenberg, and I. Thompson. Prostate rebiopsy is a poor surrogate of treatment efficacy in localized prostate cancer. *Journal of Urology*, 159(5):1606–1608, 1998.

[30] M. Norberg, L. Egevad, L. Holmberg, P. Sparen, B. J. Norlen, and C. Busch. The sextant protocol for ultrasound-guided core biopsies of the prostate underestimates the presence of cancer. *Journal of Urology*, 50(4):562–566, 1997.

[31] D. B. Plewes and W. Kucharczyk. Physics of MRI: a primer. *Journal of Magnetic Resonance Imaging*, 5(35):1038–1054, 2012.

[32] WebMD Medical Reference. Magnetic resonance imaging (MRI). http://www.webmd.com/a-to-z-guides/magnetic-resonance-imaging-mri, 2014. Accessed 27-July-2015.

[33] C. Weegenaar and G. Stoddart. Magnetic resonance imaging e-tutorial. https://www.ole.bris.ac.uk/bbcswebdav/institution/Faculty%20of%20Health%20Sciences/MB%20ChB/Radiology/MRI%20e-tutorial/page_04.htm, 2013. Accessed 28-July-2015.

[34] D. Lee and T. Carroll. *Magnetic Resonance Imaging*. Spingerlink, London, 1st edition, 2010.

[35] A. R. Padhani. Dynamic contrast-enhanced MRI in clinical oncology: current status and future directions. *Journal of Magnetic Resonance Imaging*, 16(4): 407–242, 2002.

[36] J. P. B. O'Connor, P. S. Tofts, K. A. Miles, L. M. Parkes, G. Thompson, and A. Jackson. Dynamic contrast-enhanced imaging techniques: CT and MRI. *The British Journal of Radiology*, 84(2):S112–S120, 2011.

[37] S. K. Gujar, S. Maheshwari, I. Bjorkman-Burtscher, and P. C. Sundgren. Magnetic resonance spectroscopy. *Journal of Neuro-Ophthalmology*, 25(3):217–226, 2005.

[38] P. W. Schaefer, P. E. Grant, and R. G. Gonzalezn. Diffusion-weighted MR imaging of the brain. *Radiology*, 217(2):331–345, 2000.

[39] O. Wu, R. M Dijkhuizen, and A. G. Sorensen. Multiparametric MR imaging of brain disorders. *Topics in Magnetic Resonance Imaging*, 21(2):129–138, 2010.

[40] E. J. Rummeny, P. Reimer, and W. Heindel. *MR Imaging of the body*. Georg Thieme Verlag, Stuttgart, 2nd edition, 2006.

[41] R. H. El Khouli, K. J. Macura, I. R. Kamel, M. A. Jacobs, and D. A. Bluemke. 3 Tesla dynamic contrast enhanced magnetic resonance imaging of the breast: Pharmacokinetic parameters versus conventional kinetic curve analysis. *American Journal of Roentgenology*, 197(6):1498–1505, 2011.

[42] J. Biederer, S. Mirsadraee, M. Beer, F. Molinari, C. Hintze, G. Bauman, M. Both, E. J. R. Van Beek, J. Wild, and M. Puderbach. MRI of the lung (3/3)—current applications and future perspectives. *Insights Imaging*, 3(4): 373–386, 2012.

[43] M. Puderbach, C. Hintze, S. Ley, M. Eichinger, H. U. Kauczor, and J. Biederer. MR imaging of the chest: a practical approach at 1.5t. *European Journal of Radiology*, 64(3):345–355, 2007.

[44] A. Louie. Multimodality imaging probes: Design and challenges. *Chemical Reviews*, 110(5):3146–3195, 2010.

[45] Y. Peng, Y. Jiang, C. Yang, J. B. Brown, T. Antic, I. Sethi, C. Schmid-Tannwald, M. L. Giger, S. E. Eggener, and A. Oto. Quantitative analysis of multiparametric prostate MR images: Differentiation between prostate cancer and normal tissue and correlation with gleason score-A computer-aided diagnosis development study. *Radiology*, 267(3):787–796, 2013.

[46] G. Litjens, O. Debats, J. Barentsz, N. Karssemeijer, and H. Huisman. Computer-aided detection of prostate cancer in MRI. *IEEE Transactions on Medical Imaging*, 33(5):1083–1092, 2014.

[47] M. B. Garnick, A. MacDonald, R. Glass, and S. Leighton. *Harvard Medical School 2012 Annual Report on Prostate Diseases*. Harvard Medical School, Harvard, 2012.

[48] D. T. Ginat, S. V. Destounis, R. G. Barr, B. Castaneda, J. G. Strang, and D. J. Rubens. Us elastography of breast and prostate lesions1. *Radiographics*, 29(7): 2007–2016, 2009.

[49] S. S. Taneja. Imaging in the diagnosis and management of prostate cancer. *Reviews in Urology*, 6(3):101–113, 2004.

[50] C. M. Moore, A. Ridout, and M. Emberton. The role of MRI in active surveillance of prostate cancer. *Current Opinion in Urology*, 23(3):261–267, 2003.

[51] A. P. Kirkham, M. Emberton, and C. Allen. How good is MRI at detecting and characterising cancer within the prostate? *European Urology*, 50:1163–1174, 2006.

[52] A. E. Wefer, H. Hricak, D. B. Vigneron, F. V. Coakleya, Y. Lu, J. Wefer, U. Mueller-Lisse, P. R. Carroll, and J. Kurhanewicz. Sextant localization of prostate cancer: comparison of sextant biopsy, magnetic resonance imaging and magnetic resonance spectroscopic imaging with step section histology. *Journal of Urology*, 164:400–404, 2000.

[53] M. R. Engelbrecht, G. J. Jager, R. J. Laheij, A. L. Verbeek, H. J. van Lier, and J. O. B. Jo. Local staging of prostate cancer using magnetic resonance imaging: a meta-analysis. *European Radiology*, 12:2294–2302, 2002.

[54] M. R. Engelbrecht, P. Puech, P. Colin, O. Akin, L. Lemaître, and A. Villers. Multimodality magnetic resonance imaging of prostate cancer. *Journal of Endourology*, 24(5):677–684, 2010.

[55] A. Sciarra, V. Panebianco, S. Salciccia, M. Osimani, D. Lisi, M. Ciccariello, R. Passariello, F. Di Silverio, and V. Gentile. Role of dynamic contrast-enhanced magnetic resonance (MR) imaging and proton MR spectroscopic imaging in the detection of local recurrence after radical prostatectomy for prostate cancer. *European Urology*, 54(3):589–600, 2008.

[56] H. Kajihara, Y. Hayashida, R. Murakami, K. Katahira, R. Nishimura, Y. Hamada, K. Kitani, M. Kitaoka, Y. Suzuki, M. Kitajima, T. Hirai, S. Morishita, K. Awai, and Y. Yamashita. Usefulness of diffusion-weighted imaging in the localization of prostate cancer. *International Journal of Radiation Oncology Biology Physics*, 74(2):399–403, 2009.

[57] L. K. Bittencourt, D. Hausmann, N. Sabaneeff, E. L. Gasparetto, and J. O. Barentsz. Multiparametric magnetic resonance imaging of the prostate: current concepts. *Radiologia Brasileira*, 47(5):292–300, 2014.

[58] J. L. Leake, R. L. Hardman, V. Ojili, I. Thompson, A. Shanbhogue, J. Hernandez, and J. Barentsz. Prostate MRI: Access to and current practice of prostate MRI in the united states. *Journal of the American College of Radiology*, 11(2): 156–160, 2014.

[59] C. Tempany and F. Franco. Prostate MRI: Update and current roles. *Applied Radiology*, 41(3):17–22, 2012.

[60] K. Doi. Computer-aided diagnosis in medical imaging: Historical review, current status and future potential. *Computerized medical imaging and graphics: the official journal of the Computerized Medical Imaging Society*, 31(4–5):198–211, 2007.

[61] J. T. Rothwax, A. K. George, B. J. Wood, and P. A. Pinto. Multiparametric MRI in biopsy guidance for prostate cancer: Fusion-guided. *BioMed Research International*, 2014:1–7, 2014.

[62] G. Lemaître, R. Marti, J. Freixenet, J. C. Vilanova, P. M. Walker, and F. Meriaudeau. Computer-aided detection and diagnosis for prostate cancer based on mono and multi-parametric MRI: A review. *Computers in Biology and Medicine*, 60:8–31, 2015.

[63] R. M. Summers, J. Liu, B. Rehani, P. Stafford, L. Brown, A. Louie, D. S. Barlow, D. W. Jensen, B. Cash, J. R. Choi, P. J. Pickhardt, and N. Petrick. Ctcolonography computer-aided polyp detection: Effect on radiologist observers of polyp identification by CAD on both the supine and prone scans. *Academic Radiology*, 17:948–959, 2010.

[64] S. Viswanath, B. N. Blochb, J. Chappelowa, P. Patela, N. Rofskyc, R. Lenkinskid, E. Genegad, and A. Madabhushi. Enhanced multi-protocol analysis via intelligent supervised embedding (empravise): Detecting prostate cancer on multi-parametric MRI. In *Proceedings SPIE 7963, Medical Imaging 2011: Computer-Aided Diagnosis*, 2011.

[65] P. Tiwari, S. Viswanath, J. Kurhanewicz, A. Shridhar, and A. Madabhushi. Multimodal wavelet embedding representation for data combination (maweric): integrating magnetic resonance imaging and spectroscopy for prostate cancer detection. *NMR in Biomedicine*, 25:607–619, 2012.

[66] I. Karoui, R. Fablet, J.M. Boucher, W. Pieczynski, and J. M. Augustin. Fusion of textural statistics using a similarity measure: application to texture recognition and segmentation. *Pattern Analysis Applications*, 31(3):425–434, 2008.

[67] L. Urdang. *Random House Dictionary of The English Language*. Random House, New York, 1st edition, 1968.

[68] L. G. Shapiro and G. C. Stockman. *Computer Vision*. Prentice–Hall Inc., New Jersey, 1st edition, 2001.

[69] A. A. Efros and W. T. Freeman. Image quilting for texture synthesis and transfer. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '01, pages 341–346, New York, NY, USA, 2001. ACM.

[70] G. R. Cross and A. K. Jain. Markov random field texture models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 5(1):25–39, 1983.

[71] K. Laws. *Textured Image Segmentation*. PhD thesis, University of Southern California, 1980.

[72] R. P. Taylor, A. P. Micolich, and D. Jonas. Fractal expressionism. *Physics World*, 12(10):25–28, 1999.

[73] M. Levine. *Vision in Man and Machine*. Mcgraw-Hill College, New York, 1st edition, 1985.

[74] T. Ojala, M. Pietikäinen, and D. Harwood. A comparative study of texture measures with classification based on feature distribution. *Pattern Recognition*, 29(1):51–59, 1996.

[75] T. Ojala, M. Pietikäinen, and T. Mäenpää. Gray scale and rotation invariant texture classification with local binary patterns. In *Proc. Sixth European Conference on Computer Vision (ECCV 2000), Dublin, Ireland*, pages 404–420. Springer Berlin/Heidelberg, 2000.

[76] N. Sharma, A. K. Ray, S. Sharma, K. K. Shukla, S. Pradhan, and L. M. Aggarwal. Segmentation and classification of medical images using texture-primitive features: Application of bam-type artificial neural network. *Journal of medical physics*, 33(3):119–126, 2008.

[77] C. C. Chen, J. S. DaPonte, and M. D. Fox. Fractal feature analysis and classification in medical imaging. *Medical Imaging, IEEE Transactions on*, 8(2): 133–142, 1989.

[78] D. Duda, M. Kretowski, R. Mathieu, R. de Crevoisier, and J. Bezy-Wendling. Multi-image texture analysis in classification of prostatic tissues from MRI. preliminary results. In Ewa Pietka, Jacek Kawa, and Wojciech Wieclawek, editors, *Information Technologies in Biomedicine, Volume 3*, volume 283 of *Advances in Intelligent Systems and Computing*, pages 139–150. Springer International Publishing, 2014. ISBN 978-3-319-06592-2.

[79] M. A. Sheppard and L. Shih. Efficient image texture analysis and classification for prostate ultrasound diagnosis. In *Proceedings of the 2005 IEEE Computational Systems Bioinformatics Conference - Workshops*, CSBW '05, pages 7–8, Washington, DC, USA, 2005. IEEE Computer Society. ISBN 0-7695-2442-7.

[80] C. C. Reyes Aldasoro and A. Bhalerao. Volumetric texture segmentation by discriminant feature selection and multiresolution classification. *IEEE Transactions on Medical Imaging*, 26(1):1–14, 2007.

[81] M. F. A. Fauzi and P. H. Lewis. A fully unsupervised texture segmentation algorithm. In *Proceedings of the British Machine Vision Conference (BMVC '03)*, pages 519–528, 2003.

[82] P. Brodatz. *Textures: A Photographic Album for Artists Designers*. Dover Publications,Inc., New York, 1st edition, 1966.

[83] E. A. El-Dahshana, H. M. Mohsenc, K. Revettd, and A. M. Salemf. Computer-aided diagnosis of human brain tumor through MRI: A survey and a new algorithm. *Expert Systems with Applications*, 41(11):5526–5545, 2014.

[84] M. Hanmandlu R. B. Dubey and S. K. Gupta. Region growing fro MRI brain tumor volume analysis. *Indian Journal of Science and Technology*, 2(9):26–31, 2009.

[85] S. Y. Yeo, X. Xie, I. Sazonov, and P. Nithiarasu. Segmentation of biomedical images using active contour model with robust image feature and shape prior. *International Journal for Numerical Methods in Engineering*, 30(2):232–248, 2014.

[86] J. Sachdeva, V. Kumar, I. Gupta, N. Khandelwal, and C. K. Ahuja. A novel content-based active contour model for brain tumor segmentation. *Magnetic Resonance Imaging*, 30(5):694–715, 2012.

[87] V. Pedoia and E. Binaghi. Automatic MRI 2D brain segmentation using graph searching technique. *Magnetic Resonance Imaging*, 29(9):887–904, 2003.

[88] D. Z. Chen K. Li, X. Wu and M. Sonka. Optimal surface segmentation in volumetric images—a graph-theoretic approach. *IEEE Pattern Analysis and Machine Intelligence*, 29(9):119–134, 2006.

[89] J. P. Lewis. Texture synthesis for digital painting, 1984.

[90] A. Gagalowicz and S. D. Ma. Model driven synthesis of natural textures for 3-d scenes. *Computers Graphics*, 10(2):161– 170, 1986.

[91] Li yi Wei and Marc Levoy. Fast texture synthesis using tree-structured vector quantization. pages 479–488, 2000.

[92] R. Paget and I. D. Longstaff. Texture synthesis via a noncausal nonparametric multiscale markov random field. *IEEE Transactions on Image Processing*, 7(6): 925–931, 1998.

[93] A. A. Efros and T. K. Leung. Texture synthesis by non-parametric sampling. In *The Proceedings of the Seventh IEEE International Conference on Computer Vision, 1999*, volume 2, pages 1033–1038, 1999.

[94] K. Popat and R. W. Picard. Novel cluster-based probability model for texture synthesis, classification, and compression. In *In Visual Communications and Image Processing*, pages 756–768, 1993.

[95] A. A. Efros and T. K. Leung. Texture synthesis by non-parametric sampling. In *IEEE International Conference on Computer Vision*, pages 1033–1038, Corfu, Greece, September 1999.

[96] A. A. Efros and W. T. Freeman. Image quilting for texture synthesis and transfer. In *Proceedings of Special Interest Group on Graphics and Interactive Techniques (SIGGRAPH 2001)*, pages 341–346. ACM Press / ACM SIGGRAPH, 2001.

[97] F. Bochud, C. Abbey, and M. Eckstein. Statistical texture synthesis of mammographic images with super-blob lumpy backgrounds. *Optics Express*, 4(1): 33–42, 1999.

[98] P. Rolland and H. H. Barrett. Effect of random background inhomogeneity on observer detection performance. *Journal of the Optical Society of America A*, 9 (5):649–658, 1992.

[99] C. J. Rose. *Statistical Models of Mammographic Texture and Appearance*. PhD thesis, University of Manchester, Manchester, 2005.

[100] D. Ganguly, S. Chakraborty, M. Balitanas, and T h. Kim. Medical imaging: A review. In T h. Kim, A. Stoica, and R-S. Chang, editors, *Security-Enriched Urban Computing and Smart Grid*, volume 78 of *Communications in Computer and Information Science*, pages 504–516. Springer Berlin Heidelberg, 2010.

[101] C. W. Niblack, R. Barber, W. Equitz, M. D. Flickner, E. H. Glasman, D. Petkovic, P. Yanker, and C. Faloutsos G. Taubin. Qbic project: querying images by content, using color, texture, and shape. volume 173, pages 173–187, 1993.

[102] Y. Mingqiang, K. Kidiyo, and R. Joseph. *In: Peng-Yeng Yin (ed), Pattern recognition techniques, technology and application*. I-Tech, Vienna, Austria, 2008.

[103] J. Chen and B. Tiddeman. Multi-cue facial feature detection and tracking. In *Image and Signal Processing*, volume 5099 of *Lecture Notes in Computer Science*, pages 356–367. Springer Berlin Heidelberg, 2008.

[104] J. Martin, A. Pentland, and R. Kikinis. Shape analysis of brain structures using physical and experimental modes. In *CVPR'94. IEEE Computer Society*, pages 752–755, 1994.

[105] L. Shen, R. M. Rangayyan, and J. E. Leo Desautels. Application of shape analysis to mammographic calcifications. *IEEE Transactions on medical imaging*, 13(2):263–274, 1994.

[106] R. M. Haralick. Statistical and structural approaches to texture. *Proceedings of the IEEE*, 67(5):786–804, May 1979.

[107] N. Aggarwal and R. K. Agrawal. First and second order statistics features for classification of magnetic resonance brain images. *Journal of Signal and Information Processing*, 3(2):146–153, 2012.

[108] R. M. Haralick, K. Shanmugam, and I. Dinstein. Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics*, SMC-3(6): 610–621, 1973.

[109] Z. Chen. *Mammographic Image Analysis: Risk Assessment and Microcalcification Classification Aspects*. PhD thesis, Aberystwyth University, Wales, 2013.

[110] L. de O. Bastos, P. Liatsis, and A. Conci. Automatic texture segmentation based on k-means clustering and efficient calculation of co-occurrence features. In *15th International Conference on Systems, Signals and Image Processing, 2008. IWSSIP 2008*, pages 141–144, 2008.

[111] T. K. Sarode H. B. Kekre, K. Shah and S. D. Thepade. Performance comparison of vector quantization technique kfcg with lbg, existing transforms and pca for face recognition. *Journal of Information Retrieval*, 2(1):64–71, 2000.

[112] W. Conners, M. M. Trivedi, and C. A. Harlow. Segmentation of a high-resolution urban scene using texture operators. *Computer Vision, Graphics, and Image Processing*, 25:273–310, 1984.

[113] J. K. Kim and H. Park. Statistical textural features for detection of microcalcifications in digitized mammograms. *IEEE Transactions on Medical Imaging*, 18(3):231–238, March 1999.

[114] J. S. Weszka C. R. Dyer and A. Rosenfeld. A comparative study of texture measures for terrain classification. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-6:269–285, 1976.

[115] M. M. Galloway. Texture analysis using gray level run lengths. *Computer Graphics and Image Processing*, 4:172–179, 1975.

[116] J. Zhang and T. Tan. Brief review of invariant texture analysis methods. *Pattern Recognition*, 35(3):735 – 747, 2002.

[117] T. Ojala, M. Pietikäinen, and X. Xu. Rotation-invariant texture classifcation using feature distributions. *Computer Graphics and Image Processing*, 33:43–52, 2000.

[118] R. K. Goyal, W. L. Goh, D. P. Mital, and K. L. Chan. Invariant element compactness for texture classi0cation. In *Proceedings of the Third International Conference on Automation, Robotics and Computer Vision (ICARCV'94), Singapore*, pages 1902–1096, 1994.

[119] R. K. Goyal, W. L. Goh, D. P. Mital, and K. L. Chan. A translation rotation and scale invariant texture analysis technique based on structural properties. In *Proceedings of the Third International Conference on Automation Technology (Automation, 1994), Taipei*, 1994.

[120] N. Ahuja. Dot pattern processing using voronoi neighborhoods. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, PAMI-4(3):336–343, May 1982.

[121] M. Tuceryan and A. K. Jain. Texture segmentation using voronoi polygons. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(2):211–216, 1990.

[122] T. Matsuyama, S-I. Miura, and M. Nagao. Structural analysis of natural textures by fourier transformation. *Computer Vision, Graphics, and Image Processing*, 24(3):347–362, 1983.

[123] Y. Liu, Y. Tsin, and W-C. Lin. The promise and perils of near-regular texture. *International Journal of Computer Vision*, 62(1–2):145–159, 2005.

[124] J. Serra. Theoretical bases of the leitz texture analyses system. *Leifz Science Technology Information*, 4:125–136, 1974.

[125] S. Y. Lu and K. S. Fu. A syntactic approach to texture analysis. *Computer Graphics and Image Processing*, 7:303–330, 1978.

[126] A. Materka and M. Strzelecki. Texture analysis methods: A review. Cost b11 report, Technical University of Lodz, 1998.

[127] V. Bruce, P. R. Green, and M. A. Georgeson. *Visual Perception: Physiology, Psychology, and Ecology.* Psychology Press, Exeter, UK, 1996.

[128] C. F. Stromeyer III and S. Klein. Spatial frequency channels in human vision as asymmetric (edge) mechanisms. *Vision Research*, 14(12):1409 – 1420, 1974.

[129] K.I. Laws. Rapid texture identification. In *Proceedings SPIE Conference Image Processing for Missile Guidance*, pages 1902–1096, 1994.

[130] M. Tuceryan. Moment based texture segmentation. In *Proceedings., 11th IAPR International Conference on Pattern Recognition, 1992. Vol.III. Conference C: Image, Speech and Signal Analysis*, pages 45–48, 1992.

[131] M. Unser. Local linear transforms for texture measurements. *Signal Processing*, 11(1):61–79, 1986.

[132] A. Rosenfeld and M. Thurston. Edge and curve detection for visual scene analysis. *IEEE Transactions on Computers*, C–20(5):562–569, 1971.

[133] A. Rosenfeld and J. Weszka. Picture recognition. In K. Fu(Ed.), editor, *Digital Pattern Recognition*, pages 135–166. Springer-Verlag, 1980.

[134] J. Daugman. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Optical Society of America–A*, 2(7):1160–1169, 1985.

[135] C. Lu, P. Chung, and C. Chen. Unsupervised texture segmentation via wavelet transform. *Pattern Recognition*, 30(5):729–742, 1997.

[136] A. Laine and J. Fan. Texture classification by wavelet packet signatures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(11):1186–1191, 1993.

[137] S. Mallat. Multifrequency channel decomposition of images and wavelet models. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 37(12):2091–2110, 1989.

[138] D. Marr. Chapter 2. In *Vision*, pages 54–78. Freeman, 1982.

[139] M. Varma and A. Zisserman. Classifying images of materials: Achieving viewpoint and illumination independence. In *Proceedings of the 7th European Conference on Computer Vision, Copenhagen, Denmark*, volume 3, pages 255–271. Springer-Verlag, May 2002. URL http://www.robots.ox.ac.uk/~vgg.

[140] M. Tuceryan and A. K. Jain. Handbook of pattern recognition & computer vision. chapter Texture Analysis, pages 235–276. World Scientific Publishing Co., Inc., River Edge, NJ, USA, 1993. URL http://dl.acm.org/citation.cfm?id=178866.178899.

[141] L. Kaplan and C-C. Kuo. Texture roughness analysis and synthesis via extended self similar (ess) model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(11):1043–1056, 1995.

[142] B. Chaudhuri and N. Sarkar. Texture segmentation using fractal dimension. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(1):72–77, 1995.

[143] R. Gonzalez and R. Woods. *Digital Image Processing.* Prentice Hall, Upper Saddle River, New Jersey, 3rd edition, 2008.

[144] B. B. Mandelbort. *The Fractal Geometry of Nature.* W. H. Freeman, New York, 1st edition, 1983.

[145] A. P. Pentland. Fractal-based description of natural scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):661–674, 1984.

[146] N. Jojic, B. J. Frey, and A. Kannan. Epitomic analysis of appearance and shape. In *Proceedings of the Ninth IEEE International Conference on Computer Vision - Volume 2*, ICCV '03, Washington, DC, USA, 2003. IEEE Computer Society.

[147] Y. Mu, S. Yan, Y. Liu, T. Huang, and B. Zhou. Discriminative local binary patterns for human detection in personal album. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR'08.*, pages 1–8, June 2008.

[148] T. Lindahl. Study of local binary patterns. MSc dissertation, Linköping University, Sweden, June 2007.

[149] L. Nannand A. Lumini and S. Brahnam. Local binary patterns variants as texture descriptors for medical image analysis. *Artificial Intelligence in Medicine*, 49(2):117–125, 1984.

[150] L. Nanni and A. Lumini. Regionboost learning for 2D + 3D based face recognition. *Pattern Recognition Letter*, 28(15):2063–2070, 2007.

[151] S. Liao and A. C. S. Chung. Face recognition by using elongated local binary patterns with average maximum distance gradient magnitude. In *Proceedings of the 8th Asian Conference on Computer Vision - Volume Part II*, ACCV'07, pages 672–679, Berlin, Heidelberg, 2007. Springer-Verlag.

[152] X. Tan and B. Triggs. Enhanced local texture feature sets for face recognition under difficult lighting conditions. In S. K. Zhou, W. Zhao, X. Tang, and S. Gong, editors, *Analysis and Modeling of Faces and Gestures*, volume 4778 of *Lecture Notes in Computer Science*, pages 168–182. Springer Berlin Heidelberg, 2007.

[153] A. Hafiane, G. Seetharaman, K. Palaniappan, and B. Zavidovique. Rotationally invariant hashing of median binary patterns for texture classification. In A. Campilho and M. Kamel, editors, *Image Analysis and Recognition*, volume 5112 of *Lecture Notes in Computer Science*, pages 619–629. Springer Berlin Heidelberg, 2008.

[154] B. Fröba and A. Ernst. Face detection with the modified census transform. In *Proceedings. Sixth IEEE International Conference on Automatic Face and Gesture Recognition, 2004.*, pages 91–96, May 2004.

[155] H. Jin, Q. Liu, H. Lu, and X. Tong. Face detection using improved lbp under bayesian framework. In *Third International Conference on Image and Graphics (ICIG'04)*, pages 306–309, Dec 2004.

[156] M. Heikkilä, M. Pietikäinen, and C. Schmid. Description of interest regions with center-symmetric local binary patterns. In P. Kalra and S. Peleg, editors, *Computer Vision, Graphics and Image Processing*, volume 4338 of *Lecture Notes in Computer Science*, pages 58–69. Springer Berlin Heidelberg, 2006.

[157] Y. Raja and S. Gong. Sparse multiscale local binary patterns. In *Proceedings of the British Machine Vision Conference*, pages 82.1–82.10. BMVA Press, 2006.

[158] A. Mertins. *Signal Analysis: Wavelets, Filter Banks, Time-Frequency Transforms and Applications*, chapter 6, pages 143–195. John Wiley Sons, 1999.

[159] T. Leung and J. Malik. Representing and recognizing the visual appearance of materials using three-dimensional textons. *International Journal of Computer Vision*, 43(1):29–44, June 2001.

[160] C. Schmid. Constructing models for content-based image retrieval. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001. CVPR 2001.*, volume 2, pages II–39–II–45 vol.2, 2001.

[161] M. Varma and A. Zisserman. Texture classification: Are filter banks necessary? In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 691–698, June 2003. URL http://www.robots.ox.ac.uk/~vgg.

[162] D. Gabor. Theory of communication. part 1: The analysis of information. *Journal of the Institution of Electrical Engineers - Part III: Radio and Communication Engineering*, 93(26):429–441, November 1946.

[163] D. A. Clausi and H. Deng. Design-based texture feature fusion using gabor filters and co-occurrence probabilities. *IEEE Transactions on Image Processing*, 14(7): 925–936, November 2005.

[164] C. Liu and H. Wechsler. A gabor feature classifier for face recognition. In *Proceedings. Eighth IEEE International Conference on Computer Vision, 2001. ICCV 2001.*, volume 2, pages 270–275, 2001.

[165] A. F. Costa, J. Tekli, and A. J. M. Traina. Fast fractal stack: Fractal analysis of computed tomography scans of the lung. In *Proceedings of the 2011 International ACM Workshop on Medical Multimedia Analysis and Retrieval*, MMAR '11, pages 13–18, New York, NY, USA, 2011. ACM.

[166] A. F. Costa, G. Humpire-Mamani, and A. J. M. Traina. An efficient algorithm for fractal analysis of textures. In *25th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), 2012*, pages 39–46, August 2012.

[167] P. Moreno, A. Bernardino, and J. Santos-Victor. Gabor parameter selection for local feature detection. In *Pattern Recognition and Image Analysis*, volume 3522 of *Lecture Notes in Computer Science*, pages 11–19. Springer Berlin Heidelberg, 2005.

[168] M. Varma and A. Zisserman. A statistical approach to material classification using image patch exemplars. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(11):2032–2046, 2009.

[169] B. Julesz. A theory of preattentive texture discrimination based on first-order statistics of textons. *Biological Cybernetics*, 41(2):131–138, 1981.

[170] S-C. Zhu, C e. Guo, Y. Wang, and Z. Xu. What are textons? *International Journal of Computer Vision*, 62(1–2):121–143, 2005.

[171] L. Zhang, M. Fisher, and W. Wang. Retinal vessel segmentation using gabor filter and textons. In *Proceedings of the 18th Conference on Medical Image Understanding and Analysis, MIUA'14*, pages 155–160, 2014.

[172] M. Gangeh, L. Sorensen, S. Shaker, M. Kamel, M. de Bruijne, and M. Loog. A texton-based approach for the classification of lung parenchyma in ct images. In T. Jiang, N. Navab, J. W. Pluim, and M. Viergever, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2010*, volume 6363 of *Lecture Notes in Computer Science*, pages 595–602. Springer Berlin Heidelberg, 2010.

[173] M. Tuceryan and A. K. Jain. *The Handbook of Pattern Recognition and Computer Vision*. World Scientific Publishing Co., Massachusetts, 2nd edition, 1998.

[174] H. Tamura, S. Mori, and T. Yamawaki. Texture features corresponding to visual perception. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-8(6): 460–473, 1978.

[175] The University of Auckland. Cbir: Texture features. https://www.cs.auckland.ac.nz/courses/compsci708s1c/lectures/Glect-html/topic4c708FSC.htm, 2015. Accessed 14-August-2015.

[176] M. M. Islam, Z. Dengsheng, and L. Guojun. A geometric method to compute directionality features for texture images. In *IEEE International Conference on Multimedia and Expo, 2008*, pages 1521–1524, June 2008.

[177] L. Kaplan and C-C. Kuo. Texture roughness analysis and synthesis via extended self similar (ess) model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(11):1043–1056, 1995.

[178] M. L. Giger, H-P. Chan, and J. Boone. Anniversary paper: History and status of CAD and quantitative image analysis: The role of medical physics and aapm. *Medical Physics*, 35(12):5799–5820, 2008.

[179] T. Hambrock, P. C. Vos, C. A. Hulsbergen–van de Kaa, J. O. Barentsz, and H. J. Huisman. Prostate cancer: Computer-aided diagnosis with multiparametric 3-t MR imaging—effect on observer performance. *Radiology*, 266(2):521–530, 2013.

[180] J. O. Barentsz G. Litjens and H. J. Huisman. Computer-aided detection of prostate cancer. http://www.diagnijmegen.nl/index.php/Computer-aided_detection_of_prostate_cancer, 2015. Accessed 18-August-2015.

[181] L. Li, L. Wang, Z. Feng, Z. Hu, G. Wang, X. Yuan, H. Wang, and D. Hu. Prostate cancer magnetic resonance imaging (MRI): multidisciplinary standpoint. *Quantitative Imaging in Medicine and Surgery*, 3(2), 2013.

[182] N. Karssemeijer, J. D. M. Otten, H. Rijken, and R. Holland. Computer aided detection of masses in mammograms as decision support. *The British Journal of Radiology*, 79(2):S123–S126, 2014.

[183] R. M. Summers, J. Liu, B. Rehani, P. Stafford, L. Brown, A. Louie, D. S. Barlow, D. W. Jensen, B. Cash, J. R. Choi, P. J. Pickhardt, and N. Petrick. Ct colonography computer-aided polyp detection: Effect on radiologist observers of polyp identification by CAD on both the supine and prone scans. *Academic Radiology*, 17:948–959, 2010.

[184] F. May, T. Treumann, P. Dettmar, R. Hartung, and J. Breul. Limited value of endorectal magnetic resonance imaging and transrectal ultrasonography in the staging of clinically localized prostate cancer. *British Journal of Urology International*, 87(1):66–69, 2001. ISSN 1464-410X.

[185] C. Rasch, I. Barillot, P. Remeijer, A. Touw, M. van Herk, and J. V. Lebesque. Definition of the prostate in {CT} and MRI: a multi-observer study. *International Journal of Radiation Oncology Biology Physics*, 43(1):57–66, 1999. ISSN 0360–3016.

[186] O. Ruprecht, P. Weisser, B. Bodelle, H. Ackerman, and T. J. Vogl. MRI of the prostate: Interobserver agreement compared with histopathologic outcome after radical prostatectomy. *European Journal of Radiology*, 81(3):456–460, 2010.

[187] H-P. Chan, B. Sahiner, M. A. Helvie, N. Petrick, M. A. Roubidoux, T. E. Wilson, D. D. Adler, C. Paramagul, J. S. Newman, and S. Sanjay-Gopal. Improvement of radiologists' characterization of mammographic masses by using computer-aided diagnosis: An roc study. *Radiology*, 212(3):817–827, 1999.

[188] J. C. Dean and C. C. Ilvento. Improved cancer detection using computer-aided detection with diagnostic and screening mammography: Prospective study of 104 cancers. *American Journal of Roentgenology*, 187(1):20–28, 2006.

[189] A. Madabhushi, J. Udupa, and A. Souza. Generalized scale: theory, algorithms, and application to image inhomogeneity correction. *Computer Vision Image Understanding*, 101(2):100–121, 2006.

[190] A. Madabhushi and J. K. Udupa. New methods of MR image intensity standardization via generalized scale. *Medical Physics*, 33(9):3426–3434, 2006.

[191] A. Madabhushi, J. K. Udupa, and G. Moonis. Comparing MR image intensity standardization against tissue characterizability of magnetization transfer ratio imaging. *Medical Physics*, 24(3):667–75, 2006.

[192] Y. Artan and I. S. Yetik. Prostate cancer localization using multiparametric MRI based on semi-supervised techniques with automated seed initialization. *IEEE Transactions on Information Technology in Biomedicine*, 16(6): 2986–2994, 2012.

[193] R. T. Shinohara, E. M. Sweeney, J. Goldsmith, N. S., F. J. Mateen, P. A. Calabresi, S. Jarso, D. L. Pham, D. S. Reich, and C. M. Crainiceanu. Statistical normalization techniques for magnetic resonance imaging. *NeuroImage: Clinical*, 6:9–19, 2014.

[194] P. Liu, S. Wang, B. Turkbey, K. Grant, P. Pinto, P. Choyke, B. J. Wood, and R. M. Summers. A prostate cancer computer-aided diagnosis system using multimodal magnetic resonance imaging and targeted biopsy labels. In *Proceedings of SPIE 8670, Medical Imaging 2013: Computer-Aided Diagnosis*, pages 86701G–86701G–6., 2013.

[195] Y. Artan, D. Langer, M. Haider, T. H. Van der Kwast, A. Evans, M. Wernick, and I. Yetik. Prostate cancer segmentation with multispectral MRI using cost sensitive conditional random fields. In *IEEE International Symposium on Biomedical Imaging (ISBI '09)*, pages 278–281, 2009.

[196] Y. Artan, M. A. Haider, D. L. Langer, T. H. van der Kwast, A. J. Evans, Y. Yang, M. N. Wernick, J. Trachtenberg, and I.S. Yetik. Prostate cancer localization with multispectral MRI using cost-sensitive support vector machines and conditional random fields. *IEEE Transactions on Image Processing*, 19(9): 2444–2455, 2010.

[197] S. Ozer, M. Haider, D. L. Langer, T. H. Van der Kwast, A. Evans, M. Wernick, J. Trachtenberg, and I. Yetik. Prostate cancer localization with multispectral MRI based on relevance vector machines. In *IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI '09)*, pages 73–76, 2009.

[198] S. Ozer, D. L. Langer, X. Liu, M.A. Haider, T. H. van der Kwast, A. J. Evans, Y. Yang, M. N. Wernick, and I. S. Yetik. Supervised and unsupervised methods for prostate cancer segmentation with multispectral MRI. *Medical Physics*, 37 (4):1873–1883, 2010.

[199] V. Shah, B. Turkbey, H. Mani, Y. Pang, T. Pohida, M. J. Merino, P. A. Pinto, P. L. Choyke, and M. Bernardo. Decision support system for localizing prostate cancer based on multiparametric magnetic resonance imaging. *Medical Physics*, 39(7):4093–4103, 2012.

[200] S. Viswanath, B. N. Bloch, M. Rosen, J. Chappelow, R. Toth, N. Rofsky, R. Lenkinski, E. Genega, A. Kalyanpur, and A. Madabhushi. Integrating structural and functional imaging for computer assisted detection of prostate cancer on multi-protocol in vivo 3 Tesla MRI. In *Proceedings SPIE, Conference Series*, volume 7260, 2009.

[201] L. G. Nyul, J. K. Udupa, and X. Zhang. New variants of a method of MRI scale standardization. *IEEE Transactions on Medical Imaging*, 19(2):143–150, Feb 2000.

[202] A. Buades, B. Coll, and J. M. Morel. A review of image denoising algorithms, with a new one. *Multiscale Modeling & Simulation*, 4(2):490–530, 2005.

[203] J. Mohan, V. Krishnaveni, and Y. Guo. A survey on the magnetic resonance image denoising methods. *Biomedical Signal Processing and Control*, 9:56–69, 2014.

[204] Y. Artan, M. A. Haider, D. L. Langer, and I. S. Yetik. Semi-supervised prostate cancer segmentation with multiparametric MRI. In *Proceedings International Symposium Biomedical Imaging*, pages 648–651, 2010.

[205] J. Liang and A. Bovik. Smoothing low-snr molecular images via anisotropic median-diffusion. *IEEE Transactions on Medical Imaging*, 21(4):377–384, 2002.

[206] D. Ampeliotis, A. Anonakoudi, K. Berberidis, and E. Z. Psarakis. Computer aided detection of prostate cancer using fused information from dynamic contrast enhanced and morphological magnetic resonance images. In *IEEE International Conference on Signal Processing and Communications*, pages 888–891, 2007.

[207] D. Ampeliotis, A. Anonakoudi, K. Berberidis, E. Z. Psarakis, and A. Kounoudes. A computer-aided system for the detection of prostate cancer based on magnetic resonance image analysis. In *International Symposium on Communications, Control and Signal Processing*, pages 1372–1377, 2008.

[208] R. Lopes, A. Ayache, N. Makni, P. Puech, A. Villers, S. Mordon, and N. Betrouni. Prostate cancer characterization on MR images using fractal features. *Medical Physics*, 38(1):83–95, 2011.

[209] A. Pizurica, W. Philips, I. Lemahieu, and M. Acheroy. A versatile wavelet domain noise filtration technique for medical imaging. *IEEE Transactions on Medical Imaging*, 22(3):323–331, 2003.

[210] M. Styner, C. Brechbuhler, G. Szckely, and G. Gerig. Parametric estimate of intensity inhomogeneities applied to MRI. *IEEE Transactions on Medical Imaging*, 19(3):153–165, 2000.

[211] D. Lv, X. Guo, X. Wang, J. Zhang, and J. Fang. Computerized characterization of prostate cancer by fractal analysis in MR images. *Journal of Magnetic Resonance Imaging*, 30(1):161–168, 2009.

[212] B. Zitová and J. Flusser. Image registration methods: a survey. *Image and Vision Computing*, 21(11):977–1000, 2003.

[213] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever. Mutual-information-based registration of medical images: a survey. *IEEE Transactions on Medical Imaging*, 22(8):986–1004, 2003.

[214] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu. A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, 16(5): 1190–1208, 1995.

[215] P. Viola and W. Wells III. Alignment by maximization of mutual information. *International Journal of Computer Vision*, 24(2):137–154, 1997.

[216] J. Mitra. *Multimodal image registration applied to magnetic resonance and ultrasound prostatic images*. PhD thesis, Universitat de Girona and Universit´de Bourgogne, 2012.

[217] V. Giannini, A. Vignati, S. Mazzetti, M. De Luca, C. Bracco, M. Stasi, F. Russo, E. Armando, and D. Regge. A prostate CAD system based on multiparametric analysis of dce T1-w, and dw automatically registered images. In *Proceedings of SPIE 8670, Medical Imaging 2013: Computer-Aided Diagnosis*, pages 86703E–86703E–6, 2013.

[218] S. Viswanath, B. N. Bloch, E. Genega, N. Rofsky, R. Lenkinski, J. Chappelow, R. Toth, and A. Madabhushi. A comprehensive segmentation, registration, and cancer detection scheme on 3 Tesla in vivo prostate DCE-MRI. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2008*, pages 662–669. 2008.

[219] S. Klein, U. A. van der, I. M. Lips, M. van Vulpen, M. Staring, and J. P. W. Pluim. Automatic segmentation of the prostate in 3D MR images by atlas matching using localized mutual information. *Medical Physics*, 35(4):1407–1417, 2008.

[220] MICCAI. Promise12. http://promise12.grand-challenge.org/, 2015. Accessed 18-August-2015.

[221] S. Wang, K. Burtt, B. Turkbey, P. Choyke, and R. M. Summers. Computer aided-diagnosis of prostate cancer on multiparametric MRI: A technical review of current research. *BioMed Research International*, 2014, 2014. doi: http://dx.doi.org/10.1155/2014/789561.

[222] G. Vincent, G. Guillard, and M. Bowes. Fully automatic segmentation of the prostate using active appearance models. In *Medical Image Computing and Computer Assisted Intervention (MICCAI) Grand Challenge*, Nice, France, October 2012. MICCAI.

[223] J. Zhang, M. Requardt, B. Kiefer, P. Gall, and S. K. Zhou. Region-specific hierarchical segmentation of MR prostate using discriminative learning. In *Medical Image Computing and Computer Assisted Intervention (MICCAI) Grand Challenge*, Nice, France, October 2012. MICCAI.

[224] F. Malmberg, Robin Strand, J. Kullberg, R. Nordenskjøld, and E. Bengtsson. Smart paint– a new interactive segmentation method applied to MR prostate segmentation. In *Medical Image Computing and Computer Assisted Intervention (MICCAI) Grand Challenge*, Nice, France, October 2012. MICCAI.

[225] W. van de Ven C. Hoeks S. Kerkstra B. van Ginneken G. Vincent G. Guillard N. Birbeck J. Zhang R. Strand F. Malmberg Y. Ou C. Davatzikos M. Kirschner F. Jung J. Yuan W. Qiu Q. Gao P.E. Edwards B. Maan F. van der Heijden S. Ghose J. Mitra J. Dowling D. Barratt H. Huisman A. Madabhushi G. Litjens, R. Toth. Evaluation of prostate segmentation algorithms for MRI: the promise12 challenge. *Medical Image Analysis*, 18(2):359–373, 2014.

[226] P. C. Vos, J. O. Barentsz, N. Karssemeijer, and H. J. Huisman. Automatic computer-aided detection of prostate cancer based on multiparametric magnetic resonance image analysis. *Physics in Medicine and Biology*, 57(6):1527–1542., 2012.

[227] G. J. S. Litjens, P. C. Vos, J. O. Barentsz, N. Karssemeijer, and H. J. Huisman. Automatic computer aided detection of abnormalities in multi-parametric prostate MRI. In *Proceedings of SPIE 7963, Medical Imaging 2011: Computer-Aided Diagnosis*, pages 79630T–79630T–7, 2011.

[228] G. J. S. Litjens, J. O. Barentsz, N. Karssemeijer, and H. J. Huisman. Automated computer-aided detection of prostate cancer in MR images: from a whole organ to a zone-based approach. In *Proceedings of SPIE 8315, Medical Imaging 2011: Computer-Aided Diagnosis*, pages 83150G–83150G–6, 2011.

[229] P. Tiwari, J. Kurhanewicz, and A. Madabhushi. Multi-kernel graph embedding for detection, gleason grading of prostate cancer via MRI/MRS. *IEEE Transactions on Medical Imaging*, 17(2):219–235, 2013.

[230] L. M. Bruce, C. H. Koger, and J. Li. Dimensionality reduction of hyperspectral data using discrete wavelet transform feature extraction. *IEEE Transactions on Geoscience and Remote Sensing*, 40(10):2331–2338, 2002.

[231] E. Niaf, O. Rouvière, F. Mège-Lechevallier, F. Bratan, and C. Lartizien. Computer aided diagnosis of prostate cancer in the peripheral zone using multiparametric MRI. *Physics in Medicine and Biology*, 57(12):3833–3851, 2012.

[232] J. Chappelow P. Patel N. Rofsky R. Lenkinski E. Genega A. Madabhushi S. Viswanath, B.N. Bloch. Enhanced multi-protocol analysis via intelligent supervised embedding (empravise): detecting prostate cancer on multiparametric MRI. In *Proceedings of SPIE 7963, Medical Imaging 2011: Computer-Aided Diagnosis*, 2011.

[233] I. Chan, W. Wells III, R. V. Mulkern, S. Haker, J. Zhang, K. H. Zou, S. E. Maier, and C. M. Tempany. Detection of prostate cancer by integration of line-scan diffusion, T2-mapping and T2-weighted magnetic resonance imaging; a multichannel statistical classifier. *Medical Physics*, 30(9):2390–2398, 2012.

[234] X. Liu, D. L. Langer, M. A. Haider, Y. Yang, M. N. Wernick, and I. S. Yetik. Prostate cancer segmentation with simultaneous estimation of markov random

field parameters and class. *IEEE Transactions on Medical Imaging*, 28(6):906–915, 2009.

[235] C. Chang C. Hsu and C. Lin. A practical guide to support vector classification. http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf, 2010. Accessed 20-August-2015.

[236] E. Niaf, O. Rouvière, and C. Lartizien. Computer-aided diagnosis for prostate cancer detection in the peripheral zone via multisequence MRI. In *Proceedings of SPIE 7963, Medical Imaging 2011: Computer-Aided Diagnosis*, 2011.

[237] R. Toth, S. Doyle, S. Pungavkar, A. Kalyanpur, and A. Madabhushi. A boosted ensemble scheme for accurate landmark detection for active shape models. In *SPIE Medical Imaging, vol. 7260*, 2009.

[238] B.M. Kelm, B. H. Menze, C. M. Zechmann, K. T. Baudendistel, and F. A. Hamprecht. Automated estimation of tumor probability in prostate magnetic resonance spectroscopic imaging: Pattern recognition vs quantification. *Magnetic Resonance in Medicine*, 57(1):150–159, 2007.

[239] A. Criminisi, J. Shotton, and E. Konukoglu. Decision forests: a unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning. *Foundations and Trends in Computer Graphics and Vision*, 7(2–3):81–227, 2011.

[240] T. K. Ho. The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8):832–844, 1998.

[241] B. Efron. Estimating the error rate of a prediction rule: improvement on cross-validation. *Journal of the American Statistical Association*, 78(382):316–331, 1983.

[242] H. Ito, K. Kamoi, K. Yokoyama, K. Yamada, and T. Nishimura. Visualization of prostate cancer using dynamic contrast-enhanced MRI: comparison with transrectal power doppler ultrasound. *British Journal of Radiology*, 76(909): 617–624, 2003.

[243] I. Ocak, M. Bernardo, G. Metzger, T. Barrett, P. Pinto, P. S. Albert, and P. L. Choyke. Dynamic contrast-enhanced MRI of prostate cancer at 3 t: a study

of pharmacokinetic parameters. *American Journal of Roentgenology*, 189(4): W192–W201, 2007.

[244] J. Carlsson, G. Helenius, M. G. Karlsson, O. Andren, K. Klinga-Levan, and B. Olsson. Differences in microrna expression during tumor development in the transition and peripheral zones of the prostate. *BioMed Central Cancer*, 13(1): 6093–6105, 2013.

[245] J. J. Fenton, L. Abraham, S. H. Taplin, B.M Geller, P. A. Carney, C. D'Orsi, J. G. Elmore, and W. E. Barlow. Effectiveness of computer-aided detection in community mammography practice for the breast cancer surveillance consortium. *Journal of the National Cancer Institute*, 103:1152–1161, 2011.

[246] I. Sluimer, A. Schilham, M. Prokop, and B. Ginneken. Computer analysis of computed tomography scans of the lung: a survey. *IEEE Transactions on Medical Imaging*, 25(4):385–405, 2006.

[247] H. Yoshida and A. H. Dachman. Computer-aided diagnosis for ct colonography. *Seminars in Ultrasound, CT, MRI*, 25(4):419–431, 2004.

[248] A. Savio A. Gonzalez-Pinto J. Echeveste J. M. P`rez A. Besga M. Grana, M. Termenon. Computer aided diagnosis system for alzheimer disease using brain diffusion tensor imaging features selected by pearson's correlation. *Neuroscience Letters*, 502:225–229, 2011.

[249] J. Fiot, L. D. Cohen, P. Raniga, and J. Fripp. Combining region-based and imprecise boundary-based cues for interactive medical image segmentation. *International Journal for Numerical Methods in Biomedical Engineering*, 29(9): 905–915, 2013.

[250] A. Tsai, A. Yezzi Jr., W. Wells, C. Tempany, D. Tucker, A. Fan, W. E. Grimson, and A. Willsky. A shape-based approach to the segmentation of medical imagery using level sets. *IEEE Transactions on Medical Imaging*, 22(2):103–154, 2003.

[251] J. Liu, J. Chen, X. Liu, L. Chun, and J. Tang. Mass segmentation in mammograms based on improved level set and watershed algorithm. *Advanced Intelligent Computing Theories and Applications. With Aspects of Artificial Intelligence Lecture Notes in Computer Science*, 6839:502–508, 2012.

[252] J. Shi, B. Sahiner, H. P. Chan, J. Ge, L. Hadjiiski, M. A. Helvie, A. Nees, Y. T. Wu, J. Wei, C. Zhou, Y. Zhang, and J. Cui. Characterization of mammographic

masses based on level set segmentation with new image features and patient information. *Medical Physics*, 35(1):280–290, 2008.

[253] B. Tanoori, Z. Azimifar, A. Shakibafar, and S. Katebi. Brain volumetry: an active contour model-based segmentation followed by svm-based classification. *Computers in Biology and Medicine*, 41(8):619–632, 2011.

[254] C. Hacker N. Fordyce, S. McKenna and J. Lucocq. Finding golgi stacks in electron micrographs. In *Proceedings of Medical Image Understanding and Analysis (MIUA 14)*, 2014.

[255] Y. Zhu, S. Williams, and R. Zwiggelaar. A hybrid asm approach for sparse volumetric data segmentation. *Pattern Recognition and Image Analysis*, 17(2): 252–258, 2007.

[256] Z. Chen and R. Zwiggelaar. A modified fuzzy c-means algorithm for breast tissue density segmentation in mammograms. In *10th IEEE International Conference on Information Technology and Applications in Biomedicine (ITAB), 2010*, pages 1–4, 2010.

[257] Y. J. Choi, J. K. Kim, N. Kim, K. W. Kim, E. K. Choi, and K-S. Cho. Functional MR imaging of prostate cancer. *RadioGraphics*, 27(1):63–75, 2007.

[258] N. Makni, A. Iancu, O. Colot, P. Puech, S. Mordon, and N. Betrouni. Zonal segmentation of prostate using multispectral magnetic resonance images. *Medical Physics*, 38:6093–6105, 2011.

[259] X. Liu, M. A. Haider, and S. Yetik. Automated prostate cancer localization with MRI without the need of manually extracted peripheral zone. *Medical Physics*, 36(6):2986–2994, 2011.

[260] R. E. Hendrick. *Breast MRI: Fundamentals and Technical Aspects*. Springer., New York, US, 2008.

[261] A. A. T. Bui and R. K. Taira. *Medical Imaging Informatics*. Springer., New York, US, 2010.

[262] J. O. Barentsz, J. Richenberg, R. Clements, P. Choyke, S. Verma, G. Villeirs, O. Rouviere, V. Logager, and J. J. F¨˘tterer. European society of urogenital radiology, esur prostate MR guidelines 2012. *European Radiology*, 22(4):746–757, 2012.

[263] S. E. Viswanath, N. B. Bloch, J. C. Chappelow, R. Toth, N. M. Rofsky, E. M. Genega, R. E. Lenkinski, and A. Madabhushi. Central gland and peripheral zone prostate tumors have significantly different quantitative imaging signatures on 3 Tesla endorectal, in vivo T2-weighted MR imagery. *Journal of Magnetic Resonance Imaging*, 36(1):213–224, 2012.

[264] A. M. Wink and J. B. T. M. Roerdink. Denoising functional MR images: A comparison of wavelet denoising and gaussian smoothing. *IEEE Transactions on Medical Imaging*, 23(3):374–387, 2004.

[265] K. Xiao, S. H. Ho, and Q. A. Salih. A study: segmentation of lateral ventricles in brain MRI using fuzzy c-means clustering with gaussian smoothing. In *Proceedings of the Joint Rough Set Symposium, Lecture Notes in Computer Science*, volume 4482, pages 161–170, 2007.

[266] Y. Wang, G. W. Wei, and S. Yang. Partial differential equation transform–variational formulation and fourier analysis. *International Journal for Numerical Methods in Biomedical Engineering*, 27(12):1996–2020, 2011.

[267] G. Grieg, O. Kubler, R. Kikinis, and F. A. Jolesz. Nonlinear anisotropic filtering of MRI data. *IEEE Transactions on Medical Imaging*, 11(2):221–232, 1992.

[268] R. Boyle and R. Thomas. *Computer Vision: A First Course*. Blackwell Scientific Publications, Ltd., Oxford, UK, 1988.

[269] R. Zwiggelaar, S. Astley, and C. Taylor. Detecting the central mass of a spiculated lesion using scale-orientation signatures. In N. Karssemeijer, M. Thijssen, J. Hendriks, and L. van Erning, editors, *Digital Mammography*, volume 13 of *Computational Imaging and Vision*, pages 63–70. Springer Netherlands, 1998.

[270] S. Gharge. *Segmentation of Medical Images*. PhD thesis, Narsee Monjee Institute of Management Studies, 2013.

[271] J. C. Russ. *The Image Processing Handbook*. CRC Press., Florida, UK, 4th edition, 2002.

[272] M. Bydder, D. J. Larkman, and J. V. Hajnal. Combination of signals from array coils using image-based estimation of coil sensitivity profiles. *Magnetic Resonance in Medicine*, 47(3):539–548, 2002.

[273] D. Garcia. Robust smoothing of gridded data in one and higher dimensions with missing values. *Computational Statistics and Data Analysis*, 54(4):1167–1178, 2010.

[274] P. Getreuer. A survey of gaussian convolution algorithms. *Image Processing On Line*, 3:286–310, 2012.

[275] M. A. Haidekker. *Advanced Biomedical Image Analysis*. John Wiley & Son., New Jersey, US, 2010.

[276] R. Bourne. *Fundamentals of Digital Imaging in Medicine*. Springer., London, UK, 2010.

[277] R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Practice Hall., New Jersey, US, 3rd edition, 2007.

[278] J. L. Semmlow. *Biosignal and Medical Image Processing*. Marcel Dekker Inc., New York, US, 2004.

[279] R. Lasaponara and N. Masini. *Satelite Remote Sensing: A New Tool for Archeology*. Springer., New York, US, 2012.

[280] D. Q. Zhang and S. C. Chen. A novel kernelized fuzzy cc-means algorithm with application in medical image segmentation. *Artificial Intelligence in Medicine*, 32(1):37–50, 2014.

[281] J. C. Bezdek, R. Ehrlich, and W. Full. Fcm: The fuzzy c-means clustering algorithm. *Computers Geosciences*, 10(2):37–50, 1984.

[282] L. Szilágyi, Z. Benyo, S. M. Szilágyi, and H. S. Adam. MR brain image segmentation using an enhanced fuzzy c-means algorithm. In *The 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, volume 4482, pages 724–726, 2007.

[283] W. Cai, S. Chen, and D. Zhang. Fast and robust fuzzy c-means clustering algorithms incorporating local information for image segmentation. *Pattern Recognition*, 40(3):825–838, 2007.

[284] M. Yin, R. Dhir, and A. V. Parwani. Diagnostic utility of p501s (prostein) in comparison to prostate specific antigen (psa) for the detection of metastatic prostatic adenocarcinoma. *Diagnostic Pathology*, 2(41), 2007.

[285] H. Choi, E. Loyer, H. Kaur, and P. M. Silverman. *Imaging Neoplasms of the Abdomen and Pelvis*. Hamilton (ON): BC Decker., 6th edition, 2003.

[286] R. C. Gonzalez, R. E. Woods, and S. L. Eddins. *Digital image processing using MATLAB*. Pearson Education India, 2004.

[287] S. M. Han, H. J. Lee, and J-Y. Choi. Prostate cancer detection using texture and clinical features in ultrasound image. In *International Conference on Information Acquisition, 2007. ICIA '07.*, pages 547–552, 2007.

[288] R. Clements. The role of transrectal ultrasound in diagnosing prostate cancer. *Current Urology Reports*, 3(3):194–200, 2002.

[289] S. A. Reinsberg, G. S. Payne, S. F. Riches, S. Ashley, J. M. Brewster, V. A. Morgan, and N. M. deSouza. Combined use of diffusion-weighted MRI and 1H MR spectroscopy to increase accuracy in prostate cancer detection. *American Journal of Roentgenology*, 188(1):1122–1129, 2007.

[290] A. Tabesh, M. Teverovskiy, H. Y. Pang, V. P. Kumar, D. Verbel, A. Kotsianti, and O. Saidi. Multifeature prostate cancer diagnosis and gleason grading of histological images. *IEEE Transactions on Medical Imaging*, 26(10):1366–1378, 2007.

[291] Y. S. Sung, H.-J. Kwon, B. W. Park, G. Cho, C. K. Lee, K.-S. Cho, and J. K. Kim. Prostate cancer detection on dynamic contrast-enhanced MRI: Computer-aided diagnosis versus single perfusion parameter maps. *American Journal of Roentgenology*, 197(5):1122–1129, 2011.

[292] M. R. Engelbrecht, H. J. Huisman, R. J. Laheij, G. J. Jager, G. J. van Leenders, C. A. Hulsbergen-Van De Kaa, J. J. de la Rosette, J. G. Blickman, and J. O. Barentsz. Discrimination of prostate cancer from normal peripheral zone and central gland tissue by using dynamic contrast-enhanced MR imaging. *Radiology*, 229(1):248–254, 2003.

[293] R. Shimofusa, H. Fujimoto, H. Akamata, K. Motoori, S. Yamamoto, T. Ueda, and H. Ito. Discrimination of prostate cancer from normal peripheral zone and central gland tissue by using dynamic contrast-enhanced MR imaging. *Journal of Computer Assisted Tomography*, 29(2):149–153, 2005.

[294] P. Stephens. Bbc news: Prostate cancer tests miss severity in half of cases. http://www.bbc.co.uk/news/health-26970132, 2014. Accessed 6-April-2015.

[295] S.E. Viswanath, N. B. Bloch, J. C. Chappelow, R. Toth, N. M. Rofsky, E. M. Genega, R.E. Lenkinski, and A. Madabhushi. Central gland and peripheral pone prostate tumors have significantly different quantitative imaging signatures on 3 Tesla endorectal, invivo T2-weighted MR imagery. *Journal of Magnetic Resonance Imaging*, 36(1):213–224, 2012.

[296] A. Madabhushi, M. D. Feldman, D. N. Metaxas, J. Tomaszeweski, and D. Chute. Automated detection of prostatic adenocarcinoma from high-resolution ex vivo MRI. *IEEE Transaction Medical Imaging*, 24(12):1611–1625, 2005.

[297] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer-Verlag, Inc., New York, US, 2006.

[298] J. Brownlee. A tour of machine learning algorithms. http://machinelearningmastery.com/a-tour-of-machine-learning-algorithms/, 2013. Accessed 16-August-2015.

[299] S. Parfait, P.M. Walker, G. Créhange, X. Tizon, and J. Mitéran. Classification of prostate magnetic resonance spectra using support vector machine. *Biomedical Signal Processing and Control*, 7(5):499 – 508, 2012.

[300] P. C. Vos, T. Hambrock, C. A. Hulsbergen van de Kaa, J. J. Fütterer, J. O. Barentsz, and H. J. Huisman. Computerized analysis of prostate lesions in the peripheral zone using dynamic contrast enhanced MRI. *Medical Physics*, 35(3): 888–899, 2008.

[301] V. Vapnik and A. Lerner. Pattern recognition using generalized portrait method. *Automation Remote Control*, 24:774–780, 1964.

[302] J. C. Platt. *Fast training of support vector machines using sequential minimal optimization, Advances in kernel methods: support vector learning*. MIT Press, Cambridge, MA, US, 1999.

[303] L. Breiman. Random forests. *Machine Learning*, 45:5–32, 2001.

[304] T. G. Dietterich. Ensemble methods in machine learning. In *Multiple Classifier Systems*, volume 1857 of *Lecture Notes in Computer Science*, pages 1–15. Springer Berlin Heidelberg, 2000.

[305] J. R. Quinlan. *C4.5: Programs for Machine Learning*. Morgan Kaufmann, San Francisco, CA, US, 1993.

[306] Y. Freund and L. Mason. The alternating decision tree learning algorithm. In *Proceedings of the Sixteenth International Conference on Machine Learning, Bled, Slovenia*, pages 124–133, 1999.

[307] R. E. Schapire. The boosting approach to machine learning: An overview. In D. D. Denison, M. H. Hansen, C. C. Holmes, B. Mallick, and B. Yu, editors, *Nonlinear Estimation and Classification*, volume 171 of *Lecture Notes in Statistics*, pages 149–171. Springer New York, 2003.

[308] G. Holmes, B. Pfahringer, R. Kirkby, E. Frank, and M. Hall. Multiclass alternating decision trees. In T. Elomaa, H. Mannila, and H. Toivonen, editors, *Machine Learning: ECML 2002*, volume 2430 of *Lecture Notes in Computer Science*, pages 161–172. Springer Berlin Heidelberg, 2002.

[309] C. M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[310] N. Friedman, D. Geiger, and M. Goldszmidt. An essay toward solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London*, 53:370–418, 1764.

[311] S.Mazzetti, M.DeLuca, C.Bracco, A.Vignati, V.Giannini, M.Stasi, F.Russo, E. Armando, S.Agliozzo, and D.Regge. A CAD system based on multiparametric analysis for cancer prostate detection on DCE-MRI. In *Proceedings of SPIE 7963, Medical Imaging 2011:Computer-Aided Diagnosis*, pages 79633Q–79633Q–7, 2011.

[312] N. Friedman, D. Geiger, and M. Goldszmidt. Bayesian network classifiers. *Machine Learning*, 29(2–3):131–163, 1997.

[313] N. Landwehr, M. Hall, and E. Frank. Logistic model trees. *Machine Learning*, 59(1–2):161–205, 2005.

[314] J. Friedman, T. Hastie, and R. Tibshirani. Additive logistic regression: a statistical view of boosting. *Annals of Statistics*, 28(2):337–407, 2000.

[315] E. B Baum. On the capabilities of multilayer perceptrons. *Journal of Complexity*, 4(3):193–215, 1988.

[316] N. Sadawi. Understanding multi-layer perceptron (mlp). how it works. https://www.youtube.com/watch?v=zqoNVjgJqqE, 2014. Accessed 20-August-2015.

[317] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas. On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(3):226–239, 1988.

[318] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7): 629–639, 1990.

[319] M.J. Black, G. Sapiro, D. Marimont, and D. Heeger. Robust anisotropic diffusion. *IEEE Transactions on Image Processing*, 7(3):421–432, 1998.

[320] D. Lopes. Anisotropic diffusion (perona and malik). http://uk.mathworks.com/matlabcentral/fileexchange/14995-anisotropic-diffusion--perona---malik-/content/anisodiff_Perona-Malik/anisodiff2D.m, 2007. Accessed 16-August-2015.

[321] C. Tsiotsios and M. Petrou. On the choice of the parameters for anisotropic diffusion in image processing. *Pattern Recognition*, 46(5):1369–1381, 2013.

[322] S. S. Mohamed, E. F. El-Saadany, A. Fenster, D. B. Downey, and K. Rizkalla. Region of interest identification in prostate trus images based on gabor filter. In *IEEE 46th Midwest Symposium on Circuits and Systems*, volume 1, pages 415–419, 2003.

[323] D. J. A. Margolis. Multiparametric MRI for localized prostate cancer: lesion detection and staging. *BioMed Research International*, 2014:1–11, 2014.

[324] O. Aki, E. Sala, C. S. Moskowitz, K. Kuroiwa, N. M. Ishill, D. Pucar, P. T. Scardino, and H. Hricak. Transition zone prostate cancers: features, detection, localization, and staging at endorectal MR imaging. *Radiology*, 2014:784–792, 2006.

[325] C. Brossner, A. Winterholer, M. Roehlich, E. Dlouhy-Schutz, V. Serra, M. Sonnleithner, K. H. Grubmuller, K. Pummer, and E. Schuster. Distribution of prostate carcinoma foci within the peripheral zone: analysis of 8,062 prostate biopsy cores. *World Journal of Urology*, 21(3):163–166, 2003.

[326] P. Howarth and S. Ruger. Evaluation of texture features for content-based image retrieval. *Proceedings of the International Conference on Image and Video Retrieval*, 3115:326–334, 2004.

[327] M. A. Hall. Correlation-based feature selection for discrete and numeric class machine learning. In *Proceedings 17th International Conference of Machine Learning*, pages 359–366, 2000.

[328] Z. Chen, H. Strange, A. Oliver, E. R. Denton, C. Boggis, and R. Zwiggelaar. Topological modeling and classification of mammographic microcalcification clusters. *IEEE Transactions Biomedical Engineering*, 62(4):1203–1214, 2015.

[329] J. Wang and J-D. Zucker. Solving multiple-instance problem: A lazy learning approach. In Pat Langley, editor, *17th International Conference on Machine Learning*, pages 1119–1125, 2000.

[330] D. Puig and M. A. Garcia. Determining optimal window size for texture feature extraction methods. In *Proceedings of the IX Spanish Symposium on Pattern Recognition and Image Analysis*, pages 237–242, 2001.

[331] P. Gong and P. J. Howarth. Frequency-based contextual classification and gray-level vector reduction for land-use identification. *Photogrammetric Engineering and Remote Sensing*, 58(4):423–437, 1992.

[332] J. R. Jensen. *Photogrammetric Engineering and Remote Sensing*. Prentice Hall, Inc., New Jersey, USA, 1996.

[333] M. E. Hodgson. What size window for image classification? a cognitive perspective. *Photogrammetric Engineering and Remote Sensing*, 64(8):797–807, 1998.

[334] N. Dalal and B. Triggs. Histogram of oriented gradients for human detection. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005, CVPR 2005.*, volume 1, pages 886–893, 2005.

[335] J. T. Kwak, S. Xu, B. J. Wood, B. Turkbey, P. L. Choyke, P. A. Pinto, S. Wang, and R. M. Summers. Automated prostate cancer detection using T2-weighted and high-b-value diffusion-weighted magnetic resonance imaging. *Medical Physics*, 42(5), 2015.

[336] L. van der Maaten and E. O. Postma. Texton-based texture classification. In *Proceedings of the Belgian-Dutch Artificial Intelligence Conference 2007*, pages 213–220, 2007.

[337] J. Malik, S. Belongie, J. Shi, and T. Leung. Textons, contours and regions: Cue integration in image segmentation. In *IEEE International Conference on Computer Vision, Corfu, Greece, September 1999*, pages 918–925, 1999.

[338] B. Caputo, E. Hayman, and P. Mallikarjuna. Class-specific material categorisation. In *Proceedings of the International Conference on Computer Vision, volume 2*, pages 1597–1604, 2005.

[339] O. G. Cula and K. J. Dana. 3D texture recognition using bidirectional feature histograms. *International Journal of Computer Vision*, 59(1):33–60, 2004.

[340] C. Schmid. Constructing models for content-based image retrieval. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001. CVPR 2001.*, volume 2, pages II–39–II–45, 2001.

[341] D. Gabor. Theory of communication. *Journal of the Institute of Electrical Engineers*, 93:429–457, 1946.

[342] M. Varma and A. Zisserman. Texture classification: Are filter banks necessary? In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 691—-698, 2003.

[343] J. B. MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability,Berkeley, University of California Press*, volume 1, pages 281–297, 1967.

[344] R. Xu and D. Wunsch II. Survey of clustering algorithms. *IEEE Transactions on Neural Networks*, 16(3):645–678, 2005.

[345] J. A. Hartigan and M. A. Wong. Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28(1):100–108, 1979.

[346] P. Puech, N. Betrouni, N. Makni, D. Dewalle, A. Villers, and L. Lemaitre. Computer-assisted diagnosis of prostate cancer using DCE-MRI data: design, implementation and preliminary results. *International Journal of Computer Assisted Radiology and Surgery.*, 4(1):1–10, 2009.

[347] P. Liu, S. Wang, B. Turkbey, P. C. P. Grant, K. Pinto, B. J. Wood, and R. M. Summers. A prostate cancer computer-aided diagnosis system using

multimodal magnetic resonance imaging and targerted biopsy labels. In *Proceedings of SPIE 8670, Medical Imaging 2013:Computer-Aided Diagnosis*, pages 86701G–86701G–6, 2013.

[348] T. Antic, Y. Peng, Y. Jiang, M. L. Giger, S. Eggener, and A. Oto. A study of T2-weighted MR image texture features and diffusion-weighted MR image features for computer-aided diagnosis of prostate cancer. In *Proceedings of SPIE 8670, Medical Imaging 2013:Computer-Aided Diagnosis*, pages 86701H–86701H–6, 2013.

[349] P. C. Vos, T. Hambrock, J. O. Barenstz, and H. J. Huisman. Combining T2-weighted with dynamic MR images for computerized classification of prostate lesions. In *Proceedings of SPIE 6915, Medical Imaging 2008:Computer-Aided Diagnosis*, pages 69150W–69150W–8, 2008.

[350] P. Tiwari, J. Kurhanewicz, M. Rosen, and A. Madabhushi. Semi supervised multi kernel (sesmik) graph embedding: identifying aggressive prostate cancer via magnetic resonance imaging and spectroscopy. In *Proceedings of Medical Image Computing Computer Assisted Interventions (MICCAI)*, page 666–673, 2010.

[351] E. Niaf, O. Rouvière, and F. Mège-Lechevallier. Computer-aided diagnosis for prostate cancer detection in the peripheral zone via multisequence MRI. In *Proceedings of SPIE 7963, Medical Imaging 2011: Computer-Aided Diagnosis*, pages 79633P–79633P–8, 2011.

[352] P. Tiwari, M. Rosen, G. Reed, J. Kurhanewicz, and A. Madabhushi. Spectral embedding based probabilistic boosting tree (sceptre): classifying high dimensional heterogeneous biomedical data. In *Proceedings of Medical Image Computing Computer Assisted Interventions (MICCAI)*, pages 844–851, 2009.

[353] A. J. Evans J. Trachtenberg B. C. Wilson D. L. Langer, T. H. van der Kwast and M. A. Haider. Prostate cancer detection with multi-parametric MRI: Logistic regression analysis of quantitative T2, diffusion-weighted imaging, and dynamic contrast-enhanced MRI. *Journal of Magnetic Resonance Imaging.*, 30(2):327–334, 2009.

[354] S. Viswanath, P. Tiwari, M. Rosen, and A. Madabhushi. A meta-classifier for detecting prostate cancer by quantitative integration of in vivo magnetic resonance spectroscopy and magnetic resonance imaging. In *Proceedings of SPIE 6915,*

Medical Imaging 2008: Computer-Aided Diagnosis, pages 69153D–69153D–12, 2008.

[355] P. Tiwari, M. Rosen, and A. Madabhushi. Consensus-locally linear embedding (clle): application to prostate cancer detection on magnetic resonance spectroscopy. In *Proceedings of Medical Image Computing Computer Assisted Interventions (MICCAI)*, pages 330–338, 2008.

[356] L. Matulewicz, J. F. Jansen, L. Bokacheva, H. A. Vargas, O. Akin, S. W. Fine, A. Shukla-Dave, J. A. Eastham, H. Hricak, J. A. Koutcher, and K. L. Zakian. Anatomic segmentation improves prostate cancer detection with artificial neural networks analysis of 1h magnetic resonance spectroscopic imaging. *Journal of Magnetic Resonance Imaging*, 40(6):1414–1421, 2014.